

The limits of patent protection : essays on the economics of intellectual property rights

Citation for published version (APA):

van Dijk, T. W. P. (1994). The limits of patent protection : essays on the economics of intellectual property rights. Maastricht: Datawyse / Universitaire Pers Maastricht.

Document status and date:

Published: 01/01/1994

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

The Limits of Patent Protection

Essays on the Economics of Intellectual Property Rights

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Rijksuniversiteit Limburg te Maastricht,
op gezag van de Rector Magnificus, Prof. dr. H. Philipsen,
volgens het besluit van het College van Dekanen,
in het openbaar te verdedigen
op donderdag 15 september 1994 om 14.00 uur

door

Theodorus Willem Peter van Dijk



UNIVERSITAIRE PERS MAASTRICHT

Promotores:

Prof. Dr. L.L.G. Soete

Prof. Dr. P.J.G. Van Cayseele

Beoordelingscommissie:

Prof. Dr. J.A.H. Maks (voorzitter)

Prof. Dr. R.R. De Bondt (Katholieke Universiteit Leuven)

Prof. Dr. H.J.M. Peters

CIP-DATA KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Dijk, Theodorus Willem Peter van

The limits of patent protection, essays on the economics of intellectual property rights / Theodorus Willem Peter van Dijk, - Maastricht : Universitaire Pers Maastricht. - 111.

Thesis Maastricht. - With ref. - With summary in Dutch
ISBN 90-5278-153-2

Subject Headings: patents ; industrial organization.

© 1994 Theon van Dijk, Maastricht

Druk: Datawyse Maastricht



Contents

Preface v

- 1 Introduction 1**
 - 1.1 Motivation 1
 - 1.2 Facts about Patent Protection 7
 - 1.3 Research Questions 13
 - 1.4 Methodology 14
 - 1.5 Plan of the Book 17
- 2 The Economic Theory of Patents: A Survey 19**
 - 2.1 Introduction and Some History 19
 - 2.2 The Legal and Economic Basics 23
 - 2.3 A Typical Granting Procedure 27
 - 2.4 Exploitation of the Granted Patent: Rights and Restrictions 30
 - 2.5 Economic Choices in the Granting Procedure 39
 - 2.6 Feedback to the Innovation Incentive 42
 - 2.7 Summary and Concluding Remarks 46
- 3 Patent Height and Competition in Product Improvements 49**
 - 3.1 The Model 52
 - 3.2 The Effects of Patent Height 57
 - 3.3 Patent Height and Welfare 65
 - 3.4 Conclusions 66
- 4 Price Discrimination, Dynamic Efficiency and Patent Breadth 69**
 - 4.1 Price Discrimination and Dynamic Efficiency 71
 - 4.2 An Address Model of Patent Breadth 73
 - 4.3 Economic Implications of Converging Patent Breadth in Europe 78
 - 4.4 Conclusions 80

5	On the Exploitation of Patent Protection	83
5.1	The Shape of Protection	84
5.2	Patent Exploitation by a Pure Monopolist	87
5.3	Inventing Around by Competitors: Improvement	92
5.4	Imitation	96
5.5	Improvement and Imitation Combined	101
5.6	Discussion and Conclusion	112
	Appendix A: Derivation of the Demand Curves	115
	Appendix B: Nash Price and Improvement Equilibrium	116
6	Passing the European Patent Office: Evidence from the Data Processing Industry	119
6.1	Introduction and Theoretical Backgrounds	119
6.2	Description of the Data	121
6.3	Determinants of Success Ratio Differences	123
6.4	National Waiting Times and Success Ratios	130
6.5	Conclusions	136
7	An Address Model of Patent Breadth with Two Countries	137
7.1	Related Literature	138
7.2	The Core Model	139
7.3	The Entry Game	143
7.4	The Policy Game	146
7.5	Concluding Remarks	153
8	Strengthening Worldwide Intellectual Property Protection	157
8.1	Global Patent Protection with Exogenous Innovation	161
8.2	Endogenous Innovation	166
8.3	Some Extensions of the Basic Model	169
8.4	Conclusions	176
9	Summary, Conclusions and Directions for Future Research	177
	References	185
	Samenvatting (Summary in Dutch)	197
	Curriculum Vitae	203

Preface

At the beginning of his career, Alfred Einstein worked several years at the then newly established Swiss patent office. He first started as a junior clerk but soon became a senior patent examiner. It would be overstated to say that his job at the patent office was deterministic to Einstein's later scientific theories, but the type of work - he said himself - certainly influenced his thinking and his clear way of presenting new theories. After he had left the Swiss patent office, Einstein continued to be interested in patent matters. Not only did he frequently act as an adviser for firms, he also obtained a few patents himself. Besides scientists like Einstein, with a strong interest in inventions and technological applications, patents have also attracted a lot of attention of lawyers. A number of famous patent wars have been fought out before court. To name some, the long-running conflict between Akzo and Du Pont on the superstrong aramid fibre, which will be described more extensively in chapter 1, or the famous conflict between the Holland Sweetener Company and NutraSweet about patents on the artificial sweetener aspartame.

This book is not concerned with the technical aspects of patents that Einstein was particularly interested in, nor with the legal aspects lawyers may argue about. Instead, this book deals with the economic aspects of patents. Much profit can be at stake when patents are involved. Imagine, for example, what the firm that possessed the patent on the game of monopoly must have gained from it. Unfortunately for the Parker Brothers - and this holds true for many patent-holders - they had no absolute monopoly; they faced competition from other players in the market for games.

Research is a process where, in the ideal case, positive externalities are important. Although only my name is on the cover, I have received many positive externalities from several persons for which I am most grateful. First of all, I would like to thank Luc Soete for his trust, his enthusiastic encouragement and his smooth, stimulating supervision, and particularly Patrick Van Cayseele for his

critical assessments, for showing me the rich field of Industrial Organization, and finally for his truly excellent guidance in the process of writing this thesis. Furthermore, I would like to thank Paul David, who, spread over several visits, has read each and every draft chapter and who has provided me with valuable fresh and insightful comments which were useful indeed. My special thanks go to Curtis Eaton; chapter 4 (a joint work of Patrick Van Cayseele and myself) has much profited from the comments he made when he acted as our discussant at the MERIT Conference, December 1992, and, more importantly, he delivered the essential parts for chapter 7, which became a joint work of Curtis Eaton, Patrick Van Cayseele and myself. On top of his tangible contribution, Curtis Eaton - probably without knowing it - gave me much inspiration and energy for this book. Next, I would like to thank Reinoud Joosten, Bart Verspagen and Thomas Ziesemer for the regular open-door discussions and the help they provided any time I needed them and from which I learned a lot. Finally, I would like to thank the following persons whose contributions can be traced back to the chapter level: Claude Crampes, Daniel Deneffe and Eric de Laat, for their critical remarks on chapter 3; Geert Duysters and John Hagedoorn, for providing most of the data of chapter 6; Erik Brouwer and Marga Peters, for their help in processing these data; and Ed Steinmueller, for his practical comments. By definition, positive externalities are not paid for, but I hope that in the future I can provide these persons with some externalities coming from my side.

One cannot live from intellectual externalities alone. Therefore I would like to thank some financial contributors. My major sponsor for four years has been the Foundation for the Promotion of Research in Economic Sciences (grant 450-227-014), which is part of the Netherlands Organization for Scientific Research (NWO). This organization also provided me with a grant that enabled me to visit the 3rd Summer School in Economic Theory at the Institute for Advanced Studies, June 1992, Hebrew University of Jerusalem, Israel. This grant has had a high personal rate of return. Finally, the Shell travel donation for the Crete Conference of the EARIE, September 1994, is also greatly acknowledged.

I would also like to thank all persons who make of MERIT an excellent, efficient and pleasant environment to work in; special thanks go to Wilma Coene-grachts, Mieke Donders, Corien Gijsbers (who also did a fine job of checking the English), and Silvana de Sanctis. My final thanks go to some people outside my direct working environment; my family for the solid foundation they have been providing through the years, my friends for the necessary distraction, and José for her daily support.

1.1 Motivation

It is quite remarkable that economists have devoted so much research effort to justify the patent system. After all, patents existed long before economics originated as a science. It is even more remarkable that modern economists bring forward the same arguments that were also used some five centuries ago when the first patent systems were established. During the last ten years, economists have exposed a renewed interest in the economics of ideas. A recent attempt to appraise the results of incorporating the intentional search for ideas, strongly emphasizes the characteristic of non-rivalry of ideas as being crucial to most results.¹ Non-rivalry is present if a good can be used more than once without diminishing the usefulness to others. A computer programme, for example, can be used by millions of people, but is still able to perform the same tasks for all users. The revived attention has raised the consciousness that, although ideas can cause economies to grow, a pure laissez-faire system may not perform well in providing these non-rival ideas. Institutional mechanisms can be designed to improve the invisible hand outcome. The patent system is currently mentioned as a good example of how an institutional innovation can improve welfare through collective action on non-rival goods production. Surprisingly, this same notion of non-rivalry can be found in the earliest, medieval patent systems. In latin 'patent' stands for 'open'; a patent right, an open right, is 'usable more than once', and as such a most appropriate property right to non-rival goods.

Of course, economists are able to analyze historical institutions even long after they started to exist and the newer version of this old argument is theoret-

1. See Romer (1993). Perhaps most significantly, the revived attention has influenced the field of growth theory. One of the main results of new growth theory is that the non-rival character of ideas is the driving force behind the growth of modern economies - through industries-wide increasing returns.

ically much better founded. The numerous economic analyses over time have provided us with a much better understanding of the various costs and benefits of patents. Major contributors include Machlup, Penrose, Plant, Schmookler, Arrow, and many other economists in the Industrial Organization field. Despite the considerable attention given to patents so far, some aspects still seem to be neglected in economics. The many patent cases in the juridical literature, one of which is described below, have made clear that there are limits to the monopoly power that patents provide. These limits of patent protection have hardly been analyzed by economists.

*Du Pont vs. Akzo*²

The conflict between Du Pont and Akzo was concerned with the production process of aramide, a synthetic fibre. Du Pont introduced its aramide fibre first in 1972 under the trademark of 'Kevlar'. Later, Akzo entered the market with a chemically identical fibre with the tradename of 'Twaron'. Since the base of this patent war is formed by the historical research and patent positions of both firms, some backgrounds on these positions might be useful. Du Pont has a long-standing tradition in polyamid fibre research. In the 1930s Du Pont's researchers developed a new polyamid, well known under the name of 'Nylon'. In the 1950s they developed a method for polymerising a new class of polyamids, called aromatic polyamids. This method was patented in 1958. Because the new polyamids had excellent heat-resisting properties, Du Pont's research aimed at spinning a fibre from them. In 1967 such a fibre was introduced and served as a textile fibre for fire-protective clothing. In the late 1960s research showed that a special subgroup of the new aromatic polyamids resulted not only in the most heat-resisting but also the strongest fibre known thus far. At that time, however, there was a problem in spinning the new fibre directly from the polymerisation solution. Du Pont's research was therefore mainly directed towards solving this problem. Several solutions were found and patented, but none of these solutions proved to be commercially attractive. It was not before December 1969 that Du Pont's researcher Blades improved the fibre such that it was suitable for commercial production (he developed the so-called 'air gap spinning method'). Du Pont applied for two patents for this process in 1971. The patents were granted in 1973 after several respecifications and negotiations with the US Patent Office. The patent office refused to accept the applications three times because in 1964 another firm, Monsanto Company, had applied for a related air gap spinning method. But the Blades patents were eventually granted and can be

2. For a more extensive description of the conflict see Mulder and Vergragt (1991).

considered as key patents in the protection of the fibre Kevlar, which was marketed in 1972.

Contrary to Du Pont, Akzo does not have a long tradition in polyamids. Akzo started its research on strong fibres in 1970. In the early 1970s the major goal of Akzo research was to catch up with Du Pont. By 1973 Akzo had nearly caught up and was not far from spinning its own strong fibre. There were, however, still some problems with the spinning process. Blades' patents, published in 1973, showed the solution of leading the spinning through an air gap. Because another air gap spinning process had already been described in the Monsanto patent prior to the Blades patents, the Akzo patent department advised to use parts of the process described in the Blades patents. Thus, by 1973 Akzo was also able to produce the new aromatic polyamid, which was named 'aramid' from then on.

After catching up with Du Pont, Akzo's research focused on improving the production process. As a result of this Akzo researcher Vollbracht found a new solvent for the polymerisation process in 1975. A patent application for this new solvent was filed in February 1975. The value of this alternative solvent increased enormously when it became clear that the original solvent Du Pont used was carcinogenic, while the alternative was not. Du Pont wanted a license for the new solvent, but Akzo was only prepared to grant it if a worldwide agreement was reached which enabled it to produce an aramid fibre as well. Du Pont refused to cooperate and instead tried to block the patent applications of Akzo, which caused considerable waiting lags in most countries. Probably, the reason behind Du Pont's refusal to reach an agreement was the fact that in 1975 the European synthetic fibre market collapsed. Because of financial restrictions of Akzo, Du Pont expected that Akzo had to stop the aramid project. But a loan of the Dutch Ministry of Economic Affairs prevented that. After estimating the patent positions of Akzo and Du Pont and concluding that Akzo had a sufficiently strong position, the Ministry was prepared to grant the loan. In 1978 Akzo started preparations for the building of aramid and spinning plants. As soon as it became clear that Akzo received a helping hand from the Dutch government, Du Pont started filing complaints to several courts in Europe in 1979. In reaction, Akzo contested Du Pont's patents in a US court.

In the meantime, the Akzo research department came up with another improvement in the aramid production. As a result of an improved mixing of the spinning solution, a fibre was produced which proved to be even better heat-resistant. Akzo applied for a patent (in order to obtain a stronger position in the conflict) and Du Pont opposed after its publication in 1981. The conflict escalated: Du Pont and Akzo fought each other's patents (primarily the Blades and Vollbracht patents and the latest patents on the improved mixing) in almost every industrialized country. There were victories for both (for example, Du Pont won in the US and Akzo in the Netherlands and (former) West-Germany). The conflict

was eventually ended somewhat surprisingly by a UK court. In June 1987 Akzo's aramid fibre was forbidden in the UK. Du Pont built a Kevlar spinning plant in Northern Ireland, an investment of about \$60 million, in order to strengthen its market position in Europe. In April 1988, however, the ban on Akzo's aramid in the UK was lifted and there was a possibility that Du Pont's Kevlar was going to be banned instead. To save the large investment in Northern Ireland, Du Pont finally agreed to allow Akzo to enter the international aramid fibre market in 1989.

The patent conflict between Du Pont and Akzo perfectly illustrates that the field of economics, when patents are involved, broadly intersects with the fields of technology and law. The worldwide synthetic fibre market was at stake, a new rapidly growing market offering enormous profit potential. Through intentional research, Du Pont generated polyamid innovations and developed the synthetic fibre market. The long-standing in-house research tradition in polyamids resulted in a strong patent portfolio for Du Pont. Initially, Du Pont was able to control the global aramid market and to deter entry because of its key Blades patents. However, motivated by the large profit opportunities, Akzo increased its research effort and, finally, also obtained a key patent (the Vollbracht patent). The legal war on each other's patents eventually broke the monopoly position of Du Pont and made the market a duopoly.

Have economists realized and incorporated the close connections between economics and technology and law? It took some time, but today most economists are convinced that the link between technology and economics is important. However, the relation between law and economics is largely neglected, at least in the economic patent literature. This lack of attention constitutes the basic motivation for this thesis.

Endogenous Technological Change in Economics

It took a long time for the economic literature to make the link between technology and economics. At first even Schumpeter, the forerunner among economists who pointed at the link, considered technical change as being exogenous to the economic system (see Freeman 1982, p. 212 describing the young Schumpeter's model). Later on, in his *Capitalism, Socialism and Democracy* (1942), Schumpeter adjusted his view and explained the vision of endogenous technological change, taking place in research departments of large and monopolistic firms. Only since the famous publication of Solow (1957), did most of the mainstream economists become aware of the importance of technological change for the economic system. Solow found that the productivity growth in the US over the

period 1909-1949 could be explained to a large extent by improved technologies.³ Although Solow emphasized the importance of technical change, he did not really endogenise it in his model. Kaldor (1957) was probably the first who made a theoretical economic (growth) model with endogenised technical change. In modern economic theory, the interest of growth theorists in endogenous technical change has revived and resulted in a series of 'new growth models' (Romer 1994, Grossman and Helpman 1994). But not only growth theorists have taken up the issue. At a more micro level technological change was endogenised as well. It was Arrow (1962) who first developed a theoretical model with endogenous research and development (R&D) decisions of individual firms.

The relation to patents was first made by Nordhaus (1969). Nordhaus applied Arrow's model to the study of optimal patent duration. Endogenous technical change is at the heart of his analysis; Nordhaus (p. 74) states very explicitly that the most basic justification for the institution of patents is endogenous R&D. If most inventions also occurred without intentional R&D, the basic economic argument for patents would disappear (the next chapter will discuss Nordhaus' study in more detail).

Law: a Neglected Field in Patent Economics

The link between technology and economics was hard to implement in economic science. The same seems to hold for the link between law and economics, at least in the case of patents. Economists such as Coase (1960) (and Institutional Economics in general) have explored the basic relation between law and economics, and parts of Industrial Organisation do so as well (see, for example, Scherer and Ross 1990 for the economics of antitrust legislation). But in the case of patents, the design, practice and impact of law has not received much attention in the economic science, even though patent laws and practices heavily affect economic variables. We have seen in the Du Pont-Akzo case what economic implications patents and courts can have on market structure, behaviour and performance. A large number of such cases have been discussed in the legal literature (see Cornish 1989). But despite the economic importance, the patent literature has mainly been dominated by legal experts and jurists. One could object to this view and point at, for example, the vast economic literature on patent races in Industrial

3. As a parenthesis, it is worthwhile to observe that the point made and the method used in Solow (1957) were in fact not completely new (Arrow 1994). In his article "Zur Theorie der Langfristigen Wirtschaftsentwicklung", in *Weltwirtschaftliches Archiv* in 1942, Tinbergen performed similar calculations for the period 1870-1914 for Germany, the UK, France and the US. Thus far, no or little credit has been given to Tinbergen for this (I thank Bart Verspagen for pointing this out).

Organisation. As a defense, however, one could then say that this literature is concerned with the link between economics and technology, rather than the link between economics and law. Patent race models study the economic effects of patents as incentives on the stage of R&D competition. Law is important in later stages, especially when decisions have to be made as to when to grant a patent and how to judge in cases of supposed infringement and dispute. Although some economic articles on patents have recently appeared which incorporate some ideas from law (for example, Klemperer 1990, Scotchmer and Green 1990), much remains to be done. This lack of attention constitutes the main motivation for this book.

The field of law is heavily affected by the complexity of the technical contents of patents. The implementation of patent laws becomes particularly complicated in two phases of a patent procedure. First, in the examination phase. Patent examiners must check whether applications fulfil the requirements as determined by law. Most applications filed have been carefully prepared by their applicants. The definition of the claims in applications is often the result of a delicate interplay between the inventor, a patent attorney and an economic adviser, who carefully consider all technical, legal and economic aspects of the various options when writing the claims. These claims and all kinds of technical subtleties - not of known products, but of inventions at the borders of current knowledge - must be understood and critically judged by examiners. It is thus clear that the patent office needs highly qualified persons, both technically and legally, to perform the examination phase. The technical contents of patents can heavily affect the legal practice in a later phase as well. After a patent is granted, disputes can occur because other patentholders are of the opinion that their rights are infringed or because the patentholder himself believes that his rights are violated by others. A large number of these disputes are settled outside court. The patentholder can, for example, offer the supposed infringer a license contract, or threat to file a suit against him, which may be sufficient to deter him as well. The prospect of a time-consuming and expensive court case often scares off both parties and makes a mutual settlement a more attractive alternative. Yet, there are still numerous cases where courts have to decide. The Du Pont vs. Akzo case is a good example.

To summarize, the general theme of this book is the economics of patent protection. I will not study the impact of patents on innovation incentives, where the economic analysis is strongly related to the field of technology. Such analysis could be labelled as the economics of patent *design*. Instead, the role of patents in the intersection area between law and economics is central to this book. It studies the protection that a patent provides against competitors, once the patent is granted. This field could be labelled as the economics of patent *protection*.

1.2 Facts about Patent Protection

Patent laws generally define a patent as a temporary and exclusive right on an invention. Such a legal definition must be considered as a framework which leaves much interpretation open to judges, lawyers and practical economists in firms and government service. Of course, no interpretation is involved in the legal lifetime of a patent which is fixed by law. But how far, for example, does the exclusivity of a patent reach? When exactly is a patent infringed? In theoretical economics, patents are often thought to provide perfect protection and thus absolute monopoly power. Sometimes this assumption serves the tractability of the theoretical model, but often it is believed to be realistic. If one takes a closer look at the practice of patent disputes in courts, at innovations and innovation strategies of firms and at diverse national patent systems, some facts about patent protection, which seem obvious once perceived, become clear. The list of perceived facts, which will be presented below, is not meant to be exhaustive. Only the facts which relate to the issue of patent protection are presented. These facts form the basis for the complete book: they have inspired the questions posed and the basic assumptions made in the further analyses. Some facts are strongly related to others.

(a) Most patentable inventions are patented

For a random sample of 100 US manufacturing firms, Mansfield (1986) found that about 75% of all patentable inventions are actually patented in the period 1981-1983. This percentage differs over industries. In some industries, trade secrets are more appropriate to protect inventions (for example, if technology progresses rapidly and patent procedures take too much time), or reverse engineering is so costly that patent protection is not worthwhile. Mansfield distinguished two groups of industries: one containing pharmaceuticals, chemicals, petroleum, machinery and fabricated metal products, where over 80% of the patentable inventions are patented and where patents are thus relatively important. Another, containing primary metals, electrical equipment, office equipment, instruments, motor vehicles, rubber and textiles, where patents are less important (about 60% of all inventions are patented in this group). Mansfield also observed that the proportion of patentable inventions really patented is, in most industries, independent of the impact of patents as incentives to carry out R&D. Thus, despite the fact (which he also finds for his sample) that patents do not provide strong innovation incentives (except for the pharmaceutical and chemical industry), many inventions are in fact patented. The proportion of inventions patented is in most industries positively related to firm size. Large firms make more use of patent protection than small firms. Taylor and Silberston

(1973) have obtained similar results for samples encompassing fewer industries.

(b) Inventing around a patent occurs

There is a widespread belief in economics that patents are sufficiently strong to provide perfect protection and thus absolute monopoly power. In the legal profession, however, lawyers are often confronted with the limits of patent protection. Competitors can invent around a patent with products or processes that are similar to the patented one but that are nevertheless allowed for by court. We already saw an illustration in the previous section where was described how important parts of the Blades patents of Du Pont were used by Akzo. As another illustration from real case law, consider the patent dispute about a new crane discussed in Hoyng (1990). The crane is protected by European patent #0033060. The new aspect of this crane is the flexible counterweight for the crane jib, which makes that the crane is in equilibrium in any position (see figure 1.1.a). This is obtained by a mechanical construction. Later on, a second firm entered the market with a crane that used the same idea of a flexible counterweight (see figure 1.1.b). This time the counterweight was obtained by changing the position of the counterweight by using of hydraulic pumps. The patentholder charged the other firm for infringing the patent, but the court decided that the patent was not infringed. According to the court, the functionally equivalent use of hydraulic pumps instead of bars was not obvious for an average expert in the field. Because a major part of the crane was sufficiently inventive, although the basic idea of the complete crane was similar, the second crane was not held to be too similar.

The benefits for firms that invent around existing patents are obvious. Due to spill-overs, these firms obtain information from the patentholder without paying for it. Whereas the original inventor has spent time and money to carry out R&D, imitators do not have to do this (or to a smaller degree) and still acquire the necessary information. Firms that invent around thus face a cost and time advantage relative to the original inventor. In a study of 48 new products, Mansfield, Schwartz and Wagner (1981) found that, on average, the ratio of the imitation cost to the (original) innovation cost was about 65%. On average, the ratio of the imitation time to the innovation time was about 70%. These relative R&D cost and time advantages indeed spurs inventing around: 60% of all patented and successful innovations were imitated within 4 years after introduction. The results of Mansfield, Schwartz and Wagner are confirmed in a more recent study of Levin, Klevorick, Nelson and Winter (1987). The latter make a distinction between typical and major innovations and find that the latter take more time and cost to imitate (but exact percentages as in Mansfield et al. cannot be derived from their tables).

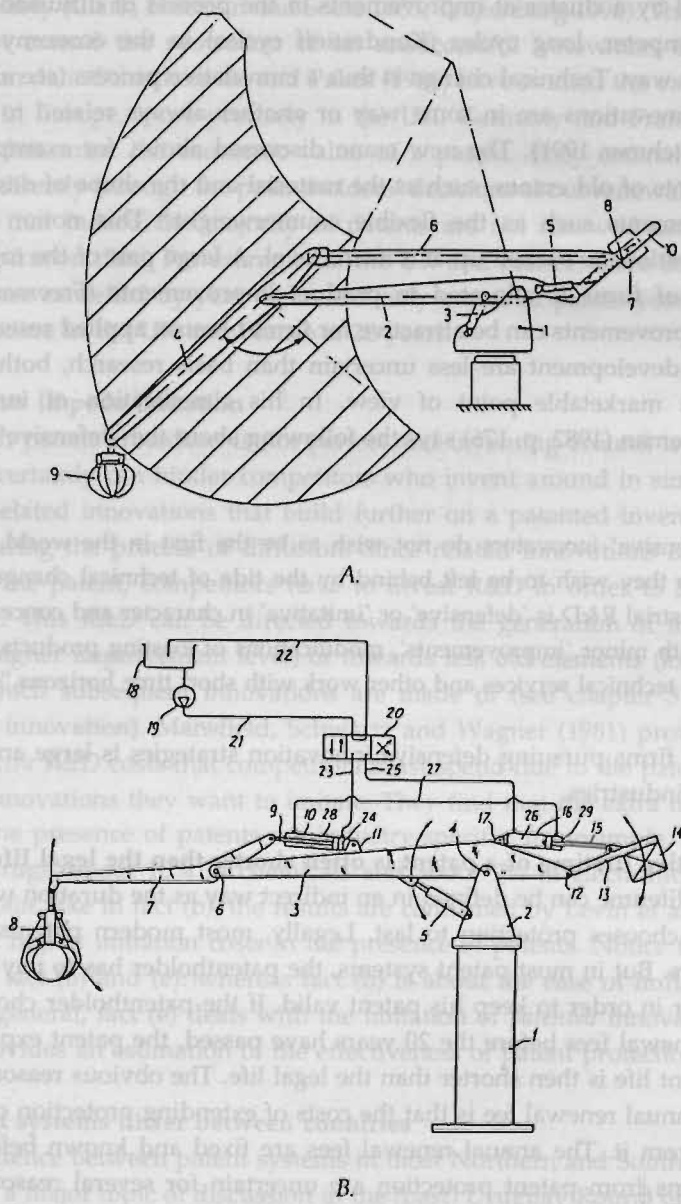


Figure 1.1 A crane patent dispute (source: Hoyng 1990)

(c) Most innovations contain elements from existing products

As Schumpeter (1942) observed, a basic invention rarely occurs isolated but is often followed by a cluster of improvements in the process of diffusion. According to Schumpeter, long cycles (Kondratieff cycles) in the economy can be explained this way. Technical change is thus a cumulative process (see also Dosi 1988). New innovations are in some way or another always related to current products (Scotchmer 1991). The new crane discussed above, for example, uses known elements of old cranes, such as the material and the shape of design, but adds new elements such as the flexible counterweight.⁴ This notion of subsequent innovations is backed up at a micro level. A large part of the total R&D expenditures of firms is allocated to product improvements (Freeman 1982). Aiming at improvements can be attractive for firms because applied research and experimental development are less uncertain than basic research, both from a technical and marketable point of view. In his classification of innovation strategies, Freeman (1982, p. 176) says the following about the 'defensive' innovation strategy:

"The 'defensive' innovators do not wish to be the first in the world, but neither do they wish to be left behind by the tide of technical change. ..., most industrial R&D is 'defensive' or 'imitative' in character and concerned mainly with minor 'improvements', modifications of existing products and processes, technical services and other work with short time horizons."

The group of firms pursuing defensive innovation strategies is large and dominant in most industries.

(d) The effective lifetime of a patent is often shorter than the legal lifetime

The effective lifetime can be defined in an indirect way as the duration which the patentholder chooses protection to last. Legally, most modern patents last for about 20 years. But in most patent systems, the patentholder has to pay renewal fees each year in order to keep his patent valid. If the patentholder chooses not to pay his renewal fees before the 20 years have passed, the patent expires. The effective patent life is then shorter than the legal life. The obvious reason for not paying the annual renewal fee is that the costs of extending protection outweigh the returns from it. The annual renewal fees are fixed and known beforehand, but the returns from patent protection are uncertain for several reasons. First, competitors may invent around the patent after some years and make the returns for the patentholder dissipate. Another reason for decreasing returns may be that

4. For more examples of related innovations, see Baker (1976).

the innovations underlying the patents are commercially not successful (any more). It is well known that the importance of patents and innovations varies considerably (Clark, Freeman and Soete 1983, Trajtenberg 1990). The patentholder may not extend the protection of an unsuccessful innovation until the legal expiration date. Schankerman and Pakes (1986) have studied the value of holding patents in Europe, more precisely in the UK, Germany and France. Their goal was to determine the economic value of a patent. They estimated the patent value indirectly through the patentholders' decisions about renewals. Their study showed that indeed the majority of patents are not renewed until the official legal expiration date. For example, in the UK and France, more than half of the patents are cancelled by year eight, and only a quarter passes year thirteen. The legal lifetime in the UK and France is 20 years.

(e) Patents impede imitation

Although patent protection is not perfect and inventing around is still possible, patents certainly can hinder competitors who invent around in similar technical fields. Related innovations that build further on a patented invention typically occur during the process of diffusion. Since related innovations can be held to infringe the patent, competitors have to invest R&D in order to be sufficiently different. This R&D can be directed towards the generation of more new elements (higher improvement level) or towards less old elements (lower imitation level) which subsequent innovations are made of (see chapter 3 for more on types of innovation). Mansfield, Schwartz and Wagner (1981) provide evidence on the *extra* R&D costs that competitors must spend due to the patent protection on the innovations they want to imitate. They find that the extra imitation costs due to the presence of patents are industry-specific: for example, about 30% in ethical drugs, about 10% in chemicals and about 7% in electronics and machinery. Again, like in fact (b), the results are confirmed by Levin et al. (1987), who also find higher imitation costs in the presence of patents. Notice the difference between fact (b) and (e): whereas fact (b) is about the ease of imitating innovations in general, fact (e) deals with the imitation of *patented* innovations and, as such, provides an estimation of the effectiveness of patent protection.

(f) Patent systems differ between countries

The difference between patent systems in most Northern and Southern countries has been a major topic of discussion in the (past) Uruguay Round of GATT negotiations (and long before, see Cooper 1973). Northern countries blamed Southern countries not to provide (effective) patent protection and wanted them to adopt

Table 1.1
Duration of patents

(a) Duration, counted from the filing date of the application:	
20 years:	Algeria; Belgium; Burundi; Chad; Denmark; Finland; France; Hungary; Israel; Italy; Monaco; Morocco; Netherlands; Nigeria; Norway; Rwanda; South Africa; Spain; Sudan; Sweden; Switzerland; United Kingdom; Zimbabwe; European Patent Convention;
16 years:	Bahamas; Jordan;
15 years:	Brazil; Barbados; Bulgaria; China; Czechoslovakia; Democratic People's Republic of Korea; Egypt; Iraq; Lebanon; Libya; Mongolia; Poland; Romania; Soviet Union; Syria; Thailand; Viet Nam;
14 years:	Malta; Mauritius;
10 years:	OAPI (African Organization); Cuba;
5, 10, 15 or 20 years*:	Iran; Tunisia;
5, 10 or 15 years**:	Turkey.
(b) Duration, counted from the date following the filing date of the application:	
20 years:	former Federal Republic of Germany (and current Germany); Luxembourg;
18 years:	former German Democratic Republic;
15 years:	Greece.
(c) Duration, counted from the publication date of the examined application:	
18 years:	Austria, but not beyond 20 years from the filing date of the application;
15 years:	Japan, but not beyond 20 years from the filing date of the application;
12 years:	Republic of Korea, but not beyond 15 years from the filing date of the application.
(d) Duration, counted from the date the complete specification is lodged:	
16 years:	Australia; Ireland; Malawi; New Zealand; Zambia;
14 years:	India.
(e) Duration, counted from the date of grant of the patent:	
17 years:	Canada; Philippines; United States of America;
16 years:	Bangladesh; Pakistan;
15 years:	Bolivia; Iceland; Malaysia; Portugal; Sri Lanka; Uruguay;
14 years:	Mexico; Trinidad and Tobago;
5, 10, or 20 years**:	Haiti;
5, 10 or 15 years*:	Argentina; Dominican Republic;
5 or 10 years*:	Venezuela;
5 years:	Colombia; Ecuador; Peru.

Source: Annex VI, p. 309-311 in Beier and Schriker (1989)

(*: depending on the applicant's request; **: the law does not appear to indicate on what the actual duration depends)

their standards of protection. But beside the obvious difference between North and South, i.e., between patent protection and no protection, more subtle differences exist within the Northern region. I will give some illustrations. First, the legal lifetime differs among Northern countries. Table 1.1 presents the patent duration of some countries. Although the legal duration is often not relevant (fact d) because the patent will not last until the end, it can matter when exactly the protection starts. The start varies considerably among countries, as does the legal lifetime. Second, the annual renewal fees differ among countries. In Germany, for example, the fee increases progressively in years, while in the UK and France the fee increases linearly and starts at a lower level (Pakes 1986). Third, the novelty requirements for patentability differ between countries. An indicator for the stringency of examination is the granting ratio of the national patent office, i.e., the proportion of applications that are granted a patent. The granting ratio in Germany was around 35% (for the period 1955-1981), in the UK 83% (for the period 1955-1981) and in France 93% (for the period 1970-1981) (see Schankerman and Pakes 1986). The final difference I will mention here is the rule in case of dispute. In Canada and the US, the first-to-invent principle is practised, which means that, if two inventors claim the same invention, the one who invented it first possibly gets it. Most other countries, on the other hand, apply the first-to-file principle. The first who files a patent application is the one who possibly obtains patent protection. This is not necessarily the first who invented.

1.3 Research Questions

The theme throughout this book is patent protection. Although each chapter will deal with one particular research question, some questions are underlying more of them. Somehow these basic questions are all inspired by the above perceived facts about patents.

The first question is on the shape of protection that a patentholder enjoys. Fact (b) says that inventing around a patent occurs. But since patents always provide some protection, inventing around cannot be unrestricted. When does a patent provide protection and when does this protection stop?

The next question is on the practice of inventing around. How do competitors technically invent around a patent? There are many possibilities. For example, given fact (c) that most innovations contain known elements and are still patentable, one manner to invent around would be by generating an improvement of a patented product. Another example is imitation. If an imitation is sufficiently dissimilar, it does not infringe the patent and is thus allowed for by court.

The third basic question is on profits. Given the shape of patent protection

and the various opportunities for inventing around, what are the profits of the patentholder? But the shape also affects the profit opportunities of the competitors who want to invent around. Fact (e) says that imitation is more costly when patents are involved.

The fourth question is on how a country designs its patent system. The profits of the patentholder and potential competitors depend on the design of patent protection. This not only holds for profits, but for the national welfare in general as well. Purely because of different isolated choices within countries, patent systems might differ between countries (fact f).

In an international context, however, the design of a domestic patent system might also depend on the foreign systems. Because of international spill-overs, countries affect each other. This might lead to national strategic behaviour in designing the patent system. The fifth and last question is then how countries interdependently design their patent systems.

Some important aspects of patenting are excluded in this book. The practice of licensing, for example, is not studied and kept outside the analysis. I refer to Kamien (1992) for a survey of theoretical models on licensing. Another aspect that is not studied here is the stage of technological competition which often precedes the stage of patenting. For an excellent review of the literature on patent races, see Reinganum (1989).

1.4 Methodology

A basic condition for the functioning of modern economies which rely to some extent on the price mechanism, is a system of well-defined property rights. Debreu's (1959) study on the general equilibrium of exchange economies uses the assumption that agents have well-defined private endowments. This assumption is necessary for the existence of a general equilibrium and its optimality in resource allocation. At first sight, the assumption of well-defined property rights seems not so stringent. Other assumptions, such as a system of markets for all goods at all places at all times for all possible states of nature, seem much more stringent. Debreu's model has proven its importance as the core of the neoclassical paradigm from which an enormous number of new theories have emerged. The economics of invention is an example of theory originally emerging from the general equilibrium framework. Arrow (1962) and Nelson (1959)⁵ have pointed out that even the assumption of well-defined property rights causes problems in

5. Arrow (1962) explains his arguments in the general equilibrium language of Debreu (1959), whereas Nelson (1959) uses similar arguments resulting from more intuitive reasoning.

the context of inventions which are the intended purpose of research and development. Due to characteristics of inventions, competitive systems do not achieve optimal resource allocation. Besides indivisibilities and uncertainty, one of the reasons that Arrow brings forth is the inappropriability of invention (which he interprets to be a broader term for knowledge). The market for inventions has some serious difficulties. In the absence of special legal protection, the inventor cannot just sell his invention on the market. A property right on knowledge can hardly be well-defined. A buyer must have a clue of what he is buying beforehand. Thus, much knowledge has to be revealed before a possible choice of purchase is made. Once bought, both the seller and the buyer have the knowledge. Nothing prevents the buyer from selling it to others. Because of this inappropriability of invention, or any product of the human mind, the market does not perform well. The institution of special property rights, such as patents, is one of the instruments to correct the market failures caused by inappropriability.

Most studies on patents leave the general equilibrium framework and take a partial equilibrium as object of study. Although some relations which are visible in general equilibrium analysis disappear, the advantage of more detailed analysis possible in an industry is often thought to be stronger. This holds for Industrial Organisation in general: generality is offered for the sake of detail. Some of the necessary, often implicit, assumptions to justify a partial equilibrium framework are: (i) that prices at other markets are fixed; and (ii) that income effects are very small (the market considered is small and price changes do not affect the consumers' budgets too much). In this book I will also pursue partial equilibrium analysis and therefore always make the above assumptions.

The principal tools used to examine the basic research questions posed in the previous section are adopted from the product differentiation literature (for an excellent short survey, see Eaton and Lipsey 1989, for a more extensive treatment, see Beath and Katsoulacos 1991 or Anderson, de Palma and Thisse 1992). There are two branches of models in the product differentiation literature. In the first, the non-address branch, consumers' preferences for differentiated goods are defined over a finite or countably infinite set of goods. The main stream within this branch uses a representative consumer approach. The seminal papers are Spence (1976) and Dixit and Stiglitz (1977). One utility function where goods enter symmetrically is representative of all consumers. The advantage of the representative consumer approach is its tractability in calculus. This explains its popularity in, for example, the new growth theory (see Grossman and Helpman 1991). There are, however, some important drawbacks in the non-address branch. For example, the fact that consumers are held to be symmetric. There is no diversity in consumers' tastes. Another example is the substitutability of goods. Since all goods are treated symmetrically, they are all equal substitutes. In real life,

however, some goods are closer substitutes than others. The address branch of product differentiation meets these criticisms.⁶ In the address branch consumers' preferences are distributed over some continuous space of parameters describing the nature of products. Consumers as well as products are characterized by addresses on this continuum. So consumers' diversity and degrees of substitutability of goods are incorporated. The price paid is a loss of calculating tractability which makes that, where the non-address branch is often used in a general equilibrium framework, the address branch is almost exclusively used in partial equilibrium analyses.

In this book I will adopt the address branch of product differentiation. Apart from the advantages mentioned above, especially in the partial equilibrium context chosen here, address models are far more appropriate to model patent dimensions. The basic notion underlying this complete book is that a patent defines a protected region on the continuum of a product (the product spectrum) which is left exclusively for the patentholder.⁷ Thus, where in basic address models firms are symmetric and have identical strategy sets, here, in the presence of a patent system, asymmetries occur between a patentholder and competitors. Given the use of product differentiation models, the analysis in this book is naturally focused on product innovations. This is in contrast with other models of patents, and microeconomic models of technical change in general, which focus primarily on process innovations (following the seminal work of Arrow 1962).

6. Archibald, Eaton and Lipsey (1986) point at some further merits of the address approach.

7. This observation, be it in a much rawer form, was already made by Nicholas Kaldor in his "Market Imperfection and Excess Capacity" (1935, p. 75, n.1): "The difference between these two types of institutional monopolies (the one which affects merely the relative costs of different producers, and the other which affects the elasticities of the demand curves for products as well) can best be elucidated by examples. A legal patent for a certain cheap process of producing ordinary window glass will not lead the consumers to differentiate between glass produced by one process or another. It will merely have the effect of imposing higher costs upon anybody who does not possess the patent. A trade-mark protecting a certain soap or medicine, however, may lead the consumers to differentiate between different soaps or medicines; and thus reduce the elasticity of demand for the products of each producer." This book extends Kaldor's interpretation of trade marks to patents.

1.5 Plan of the Book

Each chapter in this book has also been written as a separate essay,⁸ so that the reader can pick up wherever he likes. Overall though, there is a clear logic behind the chapters. First, I will present a survey of the existing theories in the economic patent literature. Next, I will present my own analyses and work from industry level to global level. I start at the micro level where dimensions of patent protection are studied, first separately and later combined. Next, models involving two and more countries are constructed. More precisely, each chapter contains the following:

Chapter 2 presents a survey of the existing literature on intellectual property. The emphasis in this survey is on theoretical models of patent protection, running from the seminal paper of Nordhaus (1969) on patent lifetime to the important article of Klemperer (1990) on patent breadth in the special patent issue of the *Rand Journal of Economics*.

Chapters 3, 4 and 5 all deal with dimensions of patent protection. Chapter 3 studies patent height, where height is defined as the strength of protection against improvements of competitors. The novelty requirements used by examiners in the patent office mainly determine the height of protection. In a model of vertical differentiation, I study the competition in product improvements between a patentholder and one competitor. Chapter 4 deals with a related but basically different dimension of protection, patent breadth. Breadth is defined as the degree of protection against imitation. The broader protection is, the less similar imitations are allowed to be. I use a model of horizontal differentiation to study the dimension of breadth. The model of patent breadth is applied to the important issue of price discrimination in competition policy. Competition policy and patent policy are always felt to conflict. Patents provide market power while competition policy often combats firms with market power. Chapter 5 combines the dimensions of height and breadth into one model. The basis of this model is a utility function which includes both vertical and horizontal differentiation. Such a model makes it possible to examine optimal strategies for inventing around patents.

Chapter 6 presents an empirical study which tests some aspects of the theo-

8. Each essay has also appeared as a MERIT Research Memorandum. A version of chapter 3 has been published in the *Tinbergen Institute Research Bulletin*. One paper (with Patrick Van Cayseele) based on chapter 4 will be published in *Technical Change and the World Economy*, edited by John Hagedoorn. Another paper based on this chapter will be published in *Economics Letters*. Chapter 7 is a joint work with Curtis Eaton and Patrick Van Cayseele.

ries developed in the previous chapters. It analyzes the patenting performance of one specific industry, the data-processing industry, in the European Patent Office. How do different innovation strategies cope with the novelty requirements of examiners in the European Patent Office?

The international setting comes in in chapters 7 and 8. First, chapter 7 presents a model of patent breadth with two countries. These countries interdependently set their patent breadths. The interdependence is caused by spill-overs which run from larger protection in one country to larger incentives in the other country. This model can deal with subtle differences in patent design between countries. Next, a model specially developed to study the point of discussion in the Uruguay Round is presented in chapter 8. Via the GATT negotiations, Northern countries want to extend their standards of protection to the South. In a simple world model, the optimal extent of patent protection worldwide is examined under various regimes.

Finally, chapter 9 summarizes the results, gives some overall concluding remarks and points at directions for further research.

The Economic Theory of Patents: A Survey

2

2.1 Introduction and Some History

The word patent originally stems from the latin 'litterae patentes', which can be translated as 'open letters'. Open here really means 'usable more than once'. Ordinary letters containing special rights were sealed on the outside. Once used, the seal was broken and the letter had no value afterwards. An open letter has the seal on the inside and the special right can thus be used more than once (see David 1993 and the references therein).

Most economic historians refer to fifteenth century Venice and seventeenth century England as the states having the oldest patent systems. Some jurists go back in time even further and mention ancient Greece as the origin of intellectual property. In the third century B.C. in Sybaris, a town in the southernmost part of current Italy, a monopoly right was granted for one year to cooks who had invented new and extraordinary recipes. The intention of this special law was that the cookery artist could exploit his invention on the one hand, and that others would be stimulated to improve the recipe of the rewarded cook on the other (Brinkhof 1988). Other historical dissertations on the origin of the patent system come up with the practice of mining in Austria in the thirteenth and fourteenth century as one of the roots of current patent law (see Kaufer 1989). The Alps were an ore-mining area with a tradition of common law concerning the granting of property rights on mining, timber and water use to those who first detected the mine. When mining became technically more complicated because the easily accessible mines were exhausted, special privileges, based on the existing mining law, were granted to those who designed devices that extended mining possibilities (such as new mechanical means to draw the water

from the lower mines).¹

Although these primitive systems all have aspects similar to modern patent systems, the first well-developed patent system existed in fifteenth century Venice. On March 19, 1474, the first general patent law was passed by the Venetian Senate (David 1993).² Another patent system, on which most modern patent systems are based, emerged in England. During the period of Elizabeth and James I, the policy of granting privileges of monopoly resulted in the Statute of Monopolies in 1623. This Statute of Monopolies is often called the Magna Charta of the right of inventors, because it was the first patent law to lay down the principle that only the first inventor should be granted a patent (Machlup and Penrose 1950).

Many countries followed the example of England and instituted patent systems.³ From the start, however, the patent system has been criticized by opponents. The criticism was strongest in the 1860s and 1870s. At that time a strong antipatent movement existed in England, Germany, and, especially, Switzerland and the Netherlands. A major explanation for the antipatent movement was the presence of a strong free-trade movement. Patent protection was intuitively associated with tariff protection. The antipatent movement was politically rather successful in weakening the patent laws. In the Netherlands the patent system was even completely abolished in July 1869, and in Switzerland the introduction was further prevented.⁴ However, from the 1870s onwards the tide changed and the victory was eventually for the advocates of the patent system. One reason for their victory was the economic crisis of 1873, which suppressed the free-trade movement and made protectionism popular. Another

1. Kitch (1977) makes a similar observation and points at the analogy between patents and mineral claims in the American West in the late nineteenth century.

2. The argument for the institution of a patent system is found in a correspondence cited in David (1993): "We have among us men of great genius, apt to invent and discover ingenious devices... Now, if provisions were made for the works and devices discovered by such persons, so that others who may see them could not build them and take the inventor's honour [sic] away, more men would then apply their genius, would discover, and would build devices of great utility to our commonwealth."

3. By the end of the eighteenth century France (1791) and the US (1793) had their patent systems. During the first half of the nineteenth century the institution of patent systems spread out. For example, Austria instituted a patent system in 1810, Russia in 1812, Prussia in 1815, Belgium and the Netherlands in 1817, Spain in 1820, Sweden in 1834, and Portugal in 1837 (Machlup and Penrose 1950).

4. See Schiff (1971) on the economic consequences of the abolishment in the Netherlands, and the absence of a patent system in Switzerland.

reason was the Paris Industrial Property Convention of 1883, where an important rule of international patenting was formulated. This rule says that nationals of each member state must be given the same treatment in other states as that given to its own nationals. If a country had no patent system, its nationals could not get protection abroad. The basic idea behind this rule was to induce countries without patent systems to institute them. Countries where the antipatent movement was strongest were thus brought to institute patent systems: Switzerland instituted a patent system in 1887 and the Netherlands re-instituted it in 1912 (Cornish 1989).

Machlup and Penrose (1950) surveyed the arguments pro and contra which were used in the patent controversy in the nineteenth century. They reported four types of arguments to justify the creation of patent rights, stressing a patent to be: (i) the natural property right in ideas; (ii) the just reward for the inventor; (iii) the best incentive to invent; and (iv) the best incentive to disclose secret information. The first argument, which was especially popular in France, says that intellectual property is not different from material property. An inventor has a natural right to his invention. The counter argument is that a natural right to an invention does not prevent others from using it (which, in fact, patents do). The second line of arguments stresses that an inventor should be compensated and rewarded for his efforts. The argument against it is that an inventor indeed needs to be rewarded, but that a patent is not necessary because rewards would also follow without. The third argument is more familiar in modern economics. Invention needs to be stimulated because of the net social gain it generates. Patents are the least expensive means to provide innovation incentives. Opponents argue that the patent system is not efficient. The costs of bureaucracy, court personnel and lawyers make the patent system too costly and unattractive. Moreover, the welfare loss due to monopoly is, of course, a major cost of the patent system as well. The final type of argument points at the disclosure of inventions, which without the patent system would have been secret to the public. Arguments brought forward against this point are, first, that inventions are often developed simultaneously by more persons. Since collective secrecy is hard to establish, disclosure would also take place without the patent system. Second, an individual inventor can hardly exploit his invention without revealing it anyway.

Nowadays these old arguments are still brought forward. They show up in the modern economic literature which this chapter will review. Other review articles on the economics of intellectual property are Penrose (1951), Vernon (1957), Machlup (1958), Kaufer (1989) and David (1993), all with major sections on historical backgrounds; Besen and Raskind (1991) and Merges and Nelson (1992), emphasizing the legal and business practices; La Manna (1992) and

Crampes and Moreaux (1993), focusing on dimensions of patent protection; and Griliches (1990), dealing with the empirical aspects of patents. The reader is referred to White et al. (1978) and Cornish (1989) for the detailed legal aspects of intellectual property.

This chapter will differentiate from these existing surveys and focus mainly on the theoretical models of patents developed in the Industrial Organization literature. After explaining the basic legal and economic aspects of intellectual property in section 2.2, the rest of the chapter is organized as follows. First, in section 2.3, I will sketch the legal and procedural aspects of patent granting in steps. The procedure as described in the European Patent Convention is typical of most modern patent systems and will be used as a guideline. The next sections will review the literature along three themes. These themes are inspired by the major stages of patentable technical change as shown in figure 2.1. The first is the stage of prepatent technological competition, where firms compete with weapons like research and development (R&D). The second stage concerns the patenting decision and the patent procedure. The final, postpatent stage is where a firm has obtained protection. The opportunities for exploitation of this protection are central to this stage.

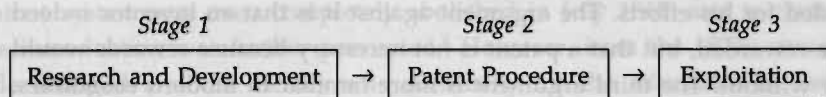


Figure 2.1 Stages in patentable technical change

I will work backwards in time, starting with the last stage. Section 2.4 will discuss models that deal with the stage of patent exploitation. How can a patentholder make his patent profitable? From the point of view of competition policy, monopoly positions are always suspicious. Patent monopolies are therefore subject to all kinds of restrictions dictated by competition policy. In addition to these legal restrictions, what are the opportunities for competitors to invent around the patent? In brief, what are the rights and restrictions of the patentholder? Next, section 2.5 will review the second stage of the granting procedure as determined by patent laws. Studies on the private and social economic choices and trade-offs relevant to this procedure are discussed. The information disclosure decision, as required by patent law, plays a major role in the literature. In addition, the requirements for patentability are studied, such as the stringency of the novelty requirements. Once the possible gains, risks and costs of patent protection are known, the first stage of research and development can be studied. The effects of

patents on the innovation incentive are central to section 2.6. Alternative ways to provide incentives are also described there. Finally, section 2.7 presents a schematic overview of the various models discussed (see table 2.2, p. 47), which may also be convenient as a guide when reading sections 2.4 through 2.6.

2.2 The Legal and Economic Basics

(a) Intellectual Property

Intellectual property can be loosely defined as the rights which can be enforced on products of the human mind. Besides patents, other important forms of intellectual property are copyright and trade marks. These three forms of intellectual property imply different rights. Patents, for example, provide an exclusive control over a technical invention for a limited time period, in general about 20 years. The exclusivity of a patent is rather strong: a patentholder can keep others from exploiting the protected product, even if they have generated it independently from the patentholder. A less strong exclusivity - applied to, for example, topographies of semiconductor chips and new plant and seed variations - allows for reverse engineering. The rightholder cannot prevent others from stripping down the invention and using it as the base for their own creation. A second form of intellectual property, copyright, provides temporary protection on literary, artistic and musical creations. Copyright can be enforced up to 50 years after the death of the creator. A copyright holder can prevent others from exactly copying or exploiting his creation. However, he cannot prevent others from exploiting their similar, independently produced creations. A third form of intellectual property, which attracts a lot of attention in the juridical literature but less in the economic literature, is the trade mark. Trade marks, such as firm names, brand names and logos, can be protected by filing them in a trade mark register. If registered, other firms are not allowed to use the same trademark. In this way, the reputation of a firm built up over time, for example through advertising, can be protected. Registered trade marks are usually protected for 7 years with the possibility of prolongation.

The patent is the intellectual property right that provides the strongest protection. The price, however, to be paid for this strong protection is the considerable effort necessary to obtain a patent. An application must be filed officially in a patent office. After time-consuming examination procedures, the patent is granted, or not. Copyright, on the other hand, is obtained almost automatically when an artistic work is finished. It suffices for authors to sign with the © mark, their name and the place of first publication. The same holds for trade marks, which only have to be registered without any examination at all.

(b) Economic Rationale

Before discussing the economics of intellectual property, it is useful to give a more detailed definition of the object that patents protect. In most juridical literature and patent laws, patents are said to protect 'inventions'. In my opinion, the term 'new technical information' would be better because it is more general and better represents the important issues. The latter term indicates that the good *information* is involved, which, as we will see below, has some special economic characteristics. This information is not about market conditions or behaviour of competitors - it is *technical* information about a product or process. In the context of patents and technological progress, particularly the *new* technical information is relevant.

The fundamental economic argument for patents is well known: because of externalities in the production of new technical information on the one hand, and because of the collective good character of information in general on the other, the social value of new technical information is often higher than the private value. The result is underprovision, even in a market with perfect competition (Arrow 1962). Patents are instruments designed to correct this market failure. They increase the private value of new technical information so that the difference between the social and private value decreases or even disappears. Less or no underprovision of new technical information in the market is the intended result.

The first element of the argument refers to the collective good character. Indeed, technical information can be considered as a collective good. A pure collective good has two characteristics (Cornes and Sandler 1986): nonrivalry and nonexcludability. Nonrivalry is present here because the consumption of information by one person does not affect the use by others. It is furthermore impossible or very costly to exclude others from consumption of information. Due to these features it is socially optimal to make the new technical information freely available (at zero price) once it is generated. The second element of the basic argument concerns the externalities in R&D. In the process of generating new technical information, externalities occur; information spills over to other firms without them paying for it. A firm can thus hardly appropriate its new technical information and the associated returns. Therefore, little or no incentive is present in private firms to generate new technical information.

The patent system is instituted to provide an incentive to carry out R&D and to exploit the collective good character as much as possible. A patentholder has the exclusive but temporary right to a piece of new technical information. The exclusivity of a patent provides monopoly power, resulting in supranormal profits, necessary to recapture the R&D costs. Only then, a firm is prepared to carry out R&D in the first place. In return for a patent, a firm does not only have

to carry out R&D, but it also has to disclose the new technical information publicly, in a way comprehensible to an expert in the field. The social benefit of information disclosure is that it prevents duplicate R&D and makes it possible for other firms to build further on it.

It is often suggested that patents are *creations* of property rights. In my opinion, it would be more precise to consider patents as *adjustments* of 'ordinary' property rights. Economic theories on the origin of the institution of ordinary property rights can also be applied to technical information. Basically, these theories (for example, Buchanan 1975) state that, without the institution of property rights, the (physical) protection of property goes along with defense and enforcement costs. If property rights are guaranteed by the state, these costs of protecting a good are no longer necessary. For this reason, the institution and enforcement of property rights by the state is beneficial to society. Property rights to inventions can be explained in the same way. An inventor could protect his invention physically and prevent others from imitating it. So why should the property right to invention be explicitly created while there exist natural rights to other property? I argue here that in order to correct the market the property right for new technical information is adjusted in such a way that market failures become less harmful. These adjustments are as follows: a patent is (i) a *temporary* and (ii) a strongly *exclusive* property right.

First, consider the exclusivity of the right. The most important aspect of a patent is not the possibility to exploit the right, but to prevent others from doing so. As mentioned above, patents also protect against firms which have independently generated a similar invention. Ordinary property rights do not have this power. The question then is: Why is this strong version of excludability needed? Two arguments can be brought forward. First, the enforcement costs that occur when proving whether or not an invention is generated independently would be very high without strong exclusivity. The enforcement costs of property rights on tangible goods may already be high and have considerable impact (see Barzel 1989, De Meza and Gould 1992), so the rights on intangible technical information are certainly hard to enforce. If the property right is made strongly exclusive, these transaction costs do not occur. An identical invention is never tolerated, independently generated or not. No proof of the patentholder is needed.

The second argument is related to the increased market and negotiation power of a patentholder who has a strongly exclusive right. I will give a simple example which highlights the point. Suppose for the moment that the lifetime of the patent is infinite and the enforcement costs are zero. An R&D lab has generated new technical information and wants to sell this. A reason for selling the information may be the inability of the lab to exploit it itself. The cost of generating the new product (the physical form of the new technical information) is c .

Other firms can also generate the new product, but they have a cost disadvantage relative to the lab; they face a cost of $c + d$. The R&D lab has thus a comparative advantage of d . The private return of the new product is v and is equal to the social return. Suppose that the R&D lab is the only supplier in the market and that there is also one demander, for example the only firm that has the right facilities to exploit the new product. In the case of ordinary property rights, the R&D lab cannot prohibit the demanding firm from generating and exploiting it itself. The maximum price that the R&D lab can therefore charge is $p = c + d$. This is the reservation price for the buyer; if the price were higher, he would be better off inventing the product himself rather than buying it. Suppose that this maximum price is indeed agreed upon. The profit for the R&D lab then is $\pi = d$, which is equal to its relative R&D cost advantage. The profit of the buyer who exploits the new product is $\pi = v - c - d$. Note that the incentive for R&D, given by the profit for the lab, is independent of the social value of the invention generated. It is completely determined by the comparative advantage in performing R&D.

If, however, the property right is as exclusive as in the case of patents, the distribution of profits changes. The reservation price of the buyer is not $p = c + d$ anymore, because the buyer is no longer permitted to generate the new product himself. The reservation price of the buyer is now $p = v$. This is what can be gained with the new product. Suppose that this reservation price becomes effective. The profit for the R&D lab then is $\pi = v - c$ and that of the buyer is $\pi = 0$. In this case, the incentive is better designed. It is the social value of the invention v that drives its generation by the R&D lab.

Summarizing, what can be learned from this simple example? The main conclusion is that in case of ordinary property rights, the incentive to innovate is solely determined by the cost advantage of the R&D lab, whereas in case of strongly exclusive rights the R&D lab is guided by the social value of the invention. The allocation of R&D means from a social point of view is thus better in the case of strongly exclusive property rights.

Besides stronger exclusivity, the second adjustment of the ordinary property right concerns the lifetime. Ordinary property rights are not temporary. Why are patents temporary? The economic reasoning goes as follows. Being a collective good, new technical information should be provided at zero price. This is, however, not possible if an incentive has to be provided. But by making the patent temporary, the new technical information will eventually be available at zero price once the patent has expired (for the question of how temporary patents should be, see Nordhaus (1969) in section 2.4).

These are the important basic economics of intellectual property and patents in particular. A vast amount of literature has emerged on the fine-tuning of these

basic arguments. Subtle parts of the patent system are more deeply analyzed. As mentioned in the Introduction, I will present this literature in a reverse chronological order. I will start by discussing the rights and restrictions once a patent is granted. Next, I will deal with economic choices in the patent-granting process. Finally, the models dealing with the impact of patents on innovation incentives will be reviewed. But first, I will describe the patent procedure of the European Patent Office, which is typical of most modern patent systems.

2.3 A Typical Granting Procedure

I will use the patent-granting procedure of the European Patent Office as a guide of exposition.⁵ Where important differences between procedures can occur, this will be mentioned. The successive steps of the procedure are as follows:

Step 1. Application and Formal Examination

A European patent application must be submitted to the European Patent Office (EPO) in Munich or its branch in The Hague. Within one month after the filing of the application, a filing fee must be paid. There are very precise prescriptions with respect to the form of application files. For example, the margins on the application files are determined by law. More importantly, the application must contain a specification of the invention, made up of two parts. First, a full description of the invention, possibly with technical drawings and schemes. Second, claims that indicate the invention where protection is looked for. An applicant and his patent attorney have to make choices when formulating these claims; broad claims are more likely to be disapproved, because claims of current patents might be infringed. If approved, however, these broad claims provide more protection.

The application remains secret up to 18 months after filing. Priority is obtained directly after filing, or, in case the same application was submitted to another patent office first, priority starts at that filing date (at least if the country of filing joined the Paris Convention). The priority date determines the point at which the application becomes state of the art.

The European patent system applies the 'first-to-file' principle; only the inventor that applies for a patent first may be granted a patent. This first-to-file system is used in most patent systems. There is another principle, used in, for example, the US and Canada, according to which not the first who files but the

5. Apart from the granting procedure, the European Patent System is not typical in some other aspects. But for the purpose of describing a patent procedure, these aspects are not important.

first who invents may obtain patent protection. The choice of the first-to-file principle or the first-to-invent principle, is particularly important in conflict situations when two firms claim the same invention.

In some countries, for example Belgium, the granting procedure stops after the first step of application. A patent is granted unconditionally and immediately. In that case the patent system is a 'registration system' comparable with trade mark systems. In countries with an 'examination system', several steps must be taken before a patent is possibly granted.

Step 2. Search

On the basis of the specification, an initial inventory of the state of the art surrounding the invention is carried out by the EPO. A fee has to be paid for this search. The state of the art is primarily estimated from past patent specifications and scientific publications. Priority rights of current applications are checked as well. The explorative search report can be used by firms to obtain insight into the state of the art in which they are active or to test whether the invention is sufficiently novel after a first inventory.

Step 3. Publication

Only after 18 months the application becomes public. These 18 months give a firm the time and opportunity to perfect the invention or to start a production process. After publication, the previously secret technical information becomes part of the state of the art. Competitors of the inventing firm then can also obtain the information. During the 18 months period of secrecy, the applicant can still decide to withdraw his application without public disclosure.

Step 4. Substantive Examination

Amendment of the application is possible between the search report and the request for examination. Requests for examination can be filed up to 6 months after the search report. The substantive examination of the patent application is guided by various requirements. A patent is only granted for an invention that (i) is new; (ii) involves an inventive step; and (iii) is industrially applicable. If these requirements are not met, a patent is not granted. Not all intellectual products can be considered inventions. For example, computer programmes and discoveries or scientific theories cannot be patented; plant and animal variations have recently been patented in the EPO, although according to the official EPO statutes this is still prohibited (Van Wijk, Cohen and Komen 1993). The novelty requirement (i) must be read as new for the public. New means: not being part of the state of technique as it is publicly known. The state of the art the moment that priority is obtained is relevant. An invention that was previously used in

secret can thus be patented. In contrast, an invention that was ever described in a scientific journal cannot.

The second requirement, the inventive step, means that an invention is not obvious for a person who is expert in the relevant technical field. Trivial developments and improvements of a new product or process are thus not patentable. If the invention is judged not to be novel enough, the possibility exists sometimes, for example in Germany, to get a utility model (or 'petty patent'). Such a utility model is similar to a regular patent but has a shorter legal lifetime (of about 6 years).

Step 5. Grant

If the application meets the requirements for patentability and validity, it passes the examination phase and a European patent is granted. To a large extent, the national patent laws of the countries that were designated in the application are relevant to the European patent. The lifetime, for example, of the European patent is determined by the national lifetimes. In order to keep the patent valid, a renewal fee has to be paid to the EPO every year. In most patent systems increase these annual renewal fees in years after granting.

Step 6. Opposition and Appeal

Up to 9 months after the grant, third parties can oppose against the granting decision of the patent office. Third parties may believe that the requirements for patentability are not fulfilled and that the patent infringes their patents. The opposition department of the EPO is charged with judging the opposition. If the opposition is held to be valid, the applicant is given the opportunity to change the specification in order to satisfy the opposition. The EPO, however, can also decide to withdraw its grant. After the opposition decision of the patent office, the initial applicant can appeal against this decision as the last possible act.

To give an impression of the decisions of the EPO and the applicants, table 2.1 gives some key figures of applications filed in the period 1978-1985. About 69% of the applications were eventually granted a patent. Of the applications that were not granted a patent, 4% was refused after formal examination, 6% was withdrawn by the applicant and 21% was deemed withdrawn (disapproved by the patent office). No opposition was filed in 82% of all grants. The opposition in the remaining 18%, was either respected, rejected or is still in process.

Table 2.1
Key figures on the EPO patent procedure (1978-1985 applications)

Procedural Status	#
Application	197,337
Refused	7,689
Withdrawn	12,244
Examination Required	177,169
Deemed Withdrawn	40,430
Granted	135,372
No Opposition	110,917

Source: Espace-Bulletin EPO

Note: Some applications are still in process. This explains the accounting differences occurring.

2.4 Exploitation of the Granted Patent: Rights and Restrictions

The restrictions on the exploitation of the patent rights and the associated monopoly position can be divided into two groups: those caused by the market and those caused by the government. The first group of restrictions stem from competitors in the market who invent around the patent. Contrary to what is commonly thought or assumed in models of, for example, patent races, patents do not provide perfect protection. The very fact that patent do not last forever but are temporary already illustrates this point. But to some extent, even during the patent lifetime, competitors can imitate and build further on the patented product. The second group of restrictions, which stem from the government, have to do with the monopoly position of the patentholder. Patent policy and competition policy are in natural conflict: while patent policy provides a legal monopoly position, competition policy often aims at suppressing it.

(a) Inventing Around by Competitors: Length, Breadth and Height of Protection

A patentholder is not perfectly protected. The protection is limited and can be described by using three dimensions: (i) the length of protection: how long does the right last?; (ii) the breadth of protection: to what extent does the patent protect against imitations? (how similar are imitations allowed to be?); and (iii) the height of protection: how novel are improvements required to be? One could

say that length indicates the possibilities for competitors to invent 'behind' the patent, breadth to invent 'aside' the patent, and height to invent 'above' the patent.

Patent Length

The historic reason in early England to set patent life to 14 years, with a possible extension of 7 years, was to protect foreign masters, who came to teach English apprentices, from competition of their apprentices once they mastered an art or skills (see David 1993 and its references). Without this protection, foreign masters would never come to England to disclose their technology and skills. Since the average time to teach apprentices was 7 years, the protection lasted for two generations of apprentices.

More formal models have been developed to study the impact of patent duration and to determine optimal lifetimes. Nordhaus (1969), the seminal work in the patent literature, describes the trade-off which a benevolent social planner faces when he chooses the lifetime of a patent. Nordhaus assumes perfect protection during the patent. The trade-off occurring is between the strength of the innovation incentive and the size of the deadweight loss. Longer patent protection induces more R&D efforts, because the expected returns are higher, and consequently improves the induced innovation, which is a larger cost reduction in Nordhaus' analysis. The welfare cost of inducing larger innovations is a longer-lasting and larger deadweight loss caused by the patent monopoly. The socially optimal length is where these two opposing forces are in balance at the margin. Nordhaus uses a framework originally developed by Arrow (1962). A major assumption underlying both studies is a monopoly in the invention market. Kamien and Schwartz (1974) and De Brock (1985) analyze the optimal patent life given a competitive invention market. Another extension is advanced by Goyal and De Laat (1991) which take more innovations into consideration.

David and Olsen (1992) point at another aspect of patent length. The longer patents last, the better learning externalities can be internalized. A monopoly with larger cumulative output, compared to more isolated firms, therefore better exploits learning effects. These positive effects of learning might offset the deadweight losses caused by the monopoly.

Although, especially in the early analyses, much emphasis is placed on the length of patent protection, there is some evidence that this dimension might be less relevant than is often thought (see also fact d in chapter 1). Empirical work by Schankerman and Pakes (1986) reveals the limited importance of *de jure* patent lifetime. Using renewal fees data, Schankerman and Pakes are able to determine the value of holding patents. As long as the expected returns from holding a patent are larger than the costs of holding the patent (determined by the annual

renewal fees), it is worthwhile to extend protection. Their data show that the effective lifetime of patents is often shorter than the legal lifetime. One possible reason may be that competitors invent around a patent during the lifetime, which makes the patent decreasingly valuable in time.

Patent Breadth

The claims in the patent specification determine the scope of patent protection. A broad protection can be obtained in two ways. First, the claims can be formulated broadly, but this might cause a number of problems. A claim can be broad to the extent that it is in the scope of protection of another patent. Moreover, if a broad protection is sought through broad claims, the applicant must have a pretty clear picture of all possible applications of the invention. Only then can the applicant protect as many potential future developments as possible. A second way to obtain broad patent protection is through the jurisdiction in patent disputes. A judge may interpret the claims broadly so that broad protection is obtained. Broad here means that the words in the claims are not taken literally, but are interpreted broadly. In the legal literature, the distinction between the 'fencepost' principle and the 'guidepost' principle⁶ is made in this context. The fencepost principle says that claims literally define the exact patent protection. According to the guidepost principle, claims only provide an indication of the protection. Equivalences, for example, of what is written down in the patent specification are also left to the exclusive use of the patentholder.

The scope of patent protection is important in infringement cases. A patentholder can suspect infringement if other firms enter the market with similar products and processes. Here and in the rest of the book, I distinguish two types of infringement. First, possible infringement from imitations which are very similar to the patented product with the exception of some minor differences. The protection against such imitations could be defined as *breadth* of protection. The second type of possible infringement stems from improvements of a patented product or process. This type will be described in the next section on *patent height*.

Within the patent breadth literature a further distinction can be made between models that define breadth for one market on the one hand, and for more markets on the other. Klemperer (1990), Waterson (1990) and Gallini (1992) belong to the first category, while Kitch (1977), Merges and Nelson (1992) and Matutes, Regibeau and Rockett (1993) belong to the other category. I will start

6. There exist some synonyms of what is labelled here as the guidepost principle. The 'signpost' system, for example, is used by Waterson (1989), or the 'equivalence' doctrine, for example, is used by Merges and Nelson (1992).

with the first category.

Since it is a seminal article and other essays in this book use similar models, I will discuss Klemperer (1990) more extensively. In a model of horizontal product differentiation, Klemperer defines patent breadth (or width as he labels it) as the distance on the product spectrum away from the patentholder's product where competing firms are allowed. All consumers are assumed to prefer, at equal prices, the patentholder's product.⁷ Consumers can differ in two aspects: first, the travel cost per unit of distance and, second, the reservation price of buying the product. A simple version of the model is illustrated in figure 2.2.

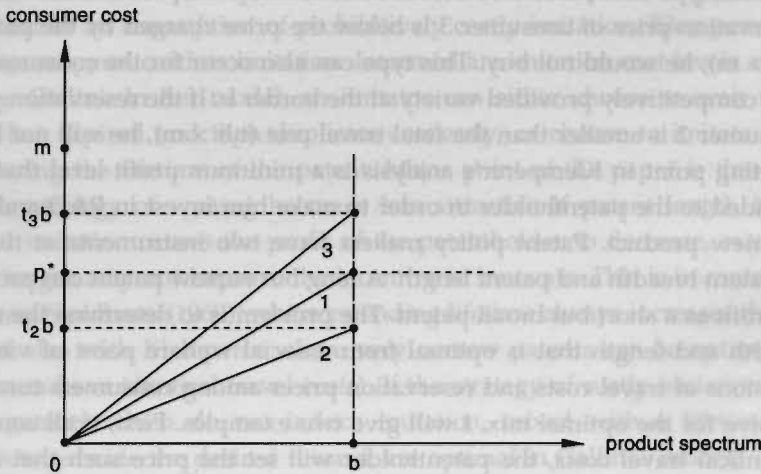


Figure 2.2 A simple illustration of the Klemperer (1990) model

The horizontal axis represents the product spectrum, with the patentholder and all consumers located at 0 and competitors located at b (given by breadth). The vertical axis represents the consumer cost per unit purchased for a consumer with reservation price m . Three consumers with identical reservation prices but

7. In the horizontal differentiation literature, it is more common to assume a (uniform) distribution of consumers over the product spectrum rather than a mass point at 0 as is done here (see Eaton and Lipsey 1989). However, if consumers have different transport costs, the analysis is very similar to one where consumers are evenly distributed over the product spectrum.

different unit travel costs t are depicted in the figure. Consumer 1 is indifferent between buying the patentholder's product at 0 with price p^* and buying the competitively provided alternative at b and incur the total travel cost of t_1b . Consumer 2 buys from the competitor because the consumer costs are lower there ($p^* > t_2b$). Consumer 3 buys from the patentholder because paying price p^* costs less than travelling to the border b ($p^* < t_3b$).

Compared to the maximum welfare which results from a competitive provision of all varieties on the spectrum, the institution of positive patent breadth induces two types of welfare losses. The first welfare loss results from consumers who shift to less preferred varieties offered at a lower price. These consumers face travel costs. Consumer 2, who incurs travel costs t_2b , is such a consumer. The second type of welfare loss occurs as consumers drop out of the market. If the reservation price of consumer 3 is below the price charged by the patentholder ($p^* > m$), he would not buy. This type can also occur for the consumers who buy the competitively provided variety at the border b . If the reservation price m for consumer 2 is smaller than the total travel cost ($t_2b > m$), he will not buy.

Starting point in Klemperer's analysis is a minimum profit level that has to be provided to the patentholder in order to make him invest in R&D and generate the new product. Patent policy makers have two instruments at their disposal: patent breadth and patent length. A long but narrow patent can provide as much profit as a short but broad patent. The problem is to determine the mixture of breadth and length that is optimal from a social welfare point of view. The distributions of travel costs and reservation prices among consumers turn out to be decisive for the optimal mix. I will give two examples. First, if all consumers face identical travel costs, the patentholder will set the price such that all consumers buy from him. In figure 2.2, suppose that all consumers have the same travel cost t_1 as consumer 1. The optimal price p^* will then be just above t_1b . No consumer will buy from the competitors. The first type of welfare loss, incurred by travelling, will thus not occur here. The second type of welfare loss, incurred by high prices, is minimized if the breadth is set very narrow and the lifetime is large enough as to provide the minimum profit required. Thus, if consumers face identical travel costs, very narrow patents are optimal. An opposite conclusion emerges if the reservation prices of all consumers are identical. Infinitely broad patent protection precludes all competition. The patentholder can then, without the need to take into account competitors, charge a price equal to the reservation price of all consumers.⁸ Neither travel costs nor deadweight losses occur. Thus, given identical reservation prices, very broad protection is optimal, mixed with a

8. This price strategy is comparable with perfect price discrimination, which is commonly known to cause no welfare loss, compared to perfect competition.

lifetime that marginally provides the minimum profit level required.

Where Klemperer assumes a fencepost patent system with an exact border of protection (b in figure 2.2), Waterson (1990) assumes a guidepost patent system with inexact coverage. Given the presence of a patent, a competitor has to decide whether or not, and where to enter the product spectrum. The potential entrant takes into account the expected cost if the patentholder takes him to court over violation of patent rights. If a court case does occur, the outcome is uncertain.⁹ The expected court costs for the imitator increase if he locates closer to the patentholder. Among the various conclusions resulting from this model, the most important is that not all inventions will be patented.

The extension of the Klemperer model by Gallini (1992) involves the cost of imitation. One could interpret Nordhaus' (1969) assumption of perfect patent protection as an assumption of prohibitively costly imitation. Klemperer, on the contrary, assumes costless imitation. Intermediate imitation cost levels are examined by Gallini. In her model, an imitator can either choose to wait until the patent has expired and then duplicate costlessly, or choose to imitate, at some cost, during the lifetime of the patent. Major point of the paper is that the first possible strategy of waiting becomes less attractive if patent length increases. Because imitation causes the profit of the patentholder to decrease, extension of patent length may not improve the incentive to innovate. This is in contrast with, of course, Nordhaus (1969), where extending lifetime always increases the incentive, but also with Klemperer where very narrow patents can be optimal (when consumers have identical travel costs). If the strategy of costly imitation is open, broad patents which do not induce imitation, mixed with a lifetime that is enough to provide the desired reward, are optimal in Gallini's model.

The models of Klemperer, Waterson and Gallini all consider the breadth of patent protection in one market. A second category of theories define patent breadth for more markets. In a mainly descriptive article Kitch (1977) exposes his prospect theory of the patent system. By prospect he means "a particular opportunity to develop a known technological possibility" (Kitch 1977, p. 266). Prospects can cover different (perhaps related) markets. Kitch argues that the process of developing prospects can efficiently be undertaken in the patent system by giving the inventor and holder of a pioneering patent sufficient control over the prospects. Most patents are granted early in the innovation process before commercialization. By providing prospects, the inventor is given more incentive to invest in developments of his invention since he does not need to be afraid of early imitation. A second advantage is that the inventor can coordinate the devel-

9. Meurer (1989) develops a model which studies the choices and possibilities for both parties when patent litigation occurs.

opment of prospects by imitators. Wasteful duplication of research can thus be avoided. Furthermore, by auctioning prospects, a prospect tends to be developed by the most efficient imitator. Kitch (1977) suggests that the US patent system has been performing the prospect function quite well by providing broad protection. Merges and Nelson (1992), in a study of patent scope decisions of US Courts, doubt as to whether the US patent scope has been broad. Moreover, contrary to Kitch, they doubt that the scope should be broad. They emphasize that if many inventors aim for the same goal, a broad patent for the lucky first one can only hinder subsequent innovations.

Although no reference is made to Kitch's prospect theory, the model of Matutes, Regibeau and Rockett (1993) could be considered as its formalization. Their starting point is a basic innovation which has various applications to be developed in the future. These possible applications, which are all known to the inventor and the potential imitators, are comparable with Kitch's prospects. In brief, the authors compare two protection regimes. The first is a regime of 'length' protection, where the patentholder has the exclusive right to introduce applications during the patent lifetime. If the patent expires, imitators can introduce applications as well. The other is a regime of 'scope' protection, where certain applications are reserved for the patentholder. Imitators can then, during the patent lifetime, introduce nonreserved applications without infringement. The regime of 'scope' protection is shown to generate higher welfare levels than the 'length' protection. The length of the period which the basic inventor chooses to wait before patenting (in which applications can be developed) and the pattern of development of subsequent innovations are taken into account when deriving this conclusion on welfare. This formal analysis tends to support the view of Kitch that inventors should be granted broad patents with control over many prospects.

Patent Height

Besides imitation, another source of possible patent infringement stems from improvement. If improvements of a patented product are introduced in the market, the incumbent patentholder is likely to lose profits. The strength of this threat for a patentholder depends primarily on the novelty requirements used by examiners in the patent office. If these requirements are more stringent and, thus, fewer improvements are approved, the current patentholder is better protected. The extent of protection against improvements is defined here as patent

height.¹⁰ So far, only few theoretical analyses have appeared on the issue of patent height. This lack of attention seems curious for two reasons. First, most industrial R&D is allocated to product improvements (see fact c in chapter 1 and Freeman 1982). Second, patenting history has shown that many patents are improvement patents (see Baker 1976).

Chapter 3 of this book presents a model where the height of protection is central. The major point there is that the competition stemming from improvements is basically different from the competition from imitations. Whereas the breadth of patent protection can be described in a model of *horizontal* differentiation, the height of protection can be described in a model of *vertical* product differentiation. The reason is that all consumers in a market are expected to prefer, at equal prices, an improvement to an older product. By definition, improvements are then vertical differentiations. They are comparable with different qualities of a product: high quality is preferred to low quality. Just like breadth can be defined on the horizontal product spectrum, height can be thought of as defining a protected region on the vertical spectrum. In a duopoly model with a patentholder and an improver, chapter 3 examines how height affects the competition in product improvements. Major conclusions are that low patents do not affect free competition (because some distance is naturally chosen), intermediate patent can benefit both the patentholder and his competitor, and high patents, by blockading entry, solely benefit the patentholder and provide him with absolute monopoly power.

A paper related to the dimension of height is Van Cayseele (1989), which discusses safety regulation in the pharmaceutical industry. From an analytical point of view, the requirements for safety imposed by health policy very much resemble the requirements for novelty imposed by patent policy. Another related paper is Scotchmer and Green (1990). Although Scotchmer and Green also deal with the stringency of the novelty requirements, their focus is not on the scope of protection but on the disclosure of information. Their paper is discussed in the next section.

(b) Conflicts with Competition Policy

The major social cost of patents, once granted, is the deadweight loss caused by the market power positions they protect. Competition policy is designed to limit or, in any case, control economic power positions. One can thus conclude that patent policy and competition policy are in natural conflict. If, as is the goal of

10. In an economic context, the term patent height was first used by Klemperer (1990, p. 127). From conversations with an EPO official, I understood that the term height is actually used within the EPO in discussions on standards of patentability.

patent policy, some reward must be given to inventors, it is clear that competition policy has to act loosely against patent monopolies. For example, if the patentholder was not allowed to set a price above marginal cost, the value of the patent would be zero. What, from a welfare point of view, would be the optimal price allowed to be set by the patentholder? Gilbert and Shapiro (1990) examine this question. Although their paper is entitled "Optimal patent length and breadth", it deals in fact more with competition policy than with patent breadth policy. Patent breadth in their model directly determines the price that a patentholder can charge. As in Klemperer (1990), a minimum profit level has to be given to a potential inventor. Two instruments are available to the government to achieve this: (i) the lifetime of protection; and (ii) the price that the patentholder is allowed to charge during the lifetime, which is directly determined by breadth. Permitting higher prices raises the patentholder's profit on the one hand, but causes the monopoly-associated deadweight losses to increase on the other. Gilbert and Shapiro find that if higher prices are increasingly costly in terms of deadweight loss, infinitely lived patents are optimal. In other words, it is better then to set lower maximum prices than shorter patent lives.

Hausman and MacKie-Mason (1988) analyze the practice of third-degree price discrimination by a patentholder. Price discrimination is often not allowed for by competition policy, not even in the case of patentholders where the rules are usually softened. The authors show that, under conditions that will be explained below, price discrimination as compared to uniform pricing raises static welfare. This result holds for price-discriminating monopolists in general, but is especially relevant for patentholders because the conditions under which it holds are particularly relevant to new products. One condition is that new markets are served under price discrimination, which are not served with uniform prices. But even without opening new markets, a Pareto improvement occurs if scale economies are present, which is typically the case for new, patented products.

Finally, Gilbert and Newberry (1982) discuss the role of patents in preemption practices. A patent creates opportunities for a firm with monopoly power to maintain this power. Indeed, a monopolist has an incentive to patent new technologies before potential competitors and thus to stay monopolist. Their point is that the extra cost of accelerated R&D can be smaller than the profit lost when a competitor enters the market. Such preemptive activity can even lead to 'sleeping' patents, i.e., patents which are neither used nor licensed. The patenting time as determined by technological competition is then earlier than the optimal introduction time, determined, among other things, by the existing capital stock of the firm. The disadvantage of the practice of preemptive patenting for social welfare is obvious. A monopoly position is sustained and, in the case of sleeping patents, new technologies, for which R&D was carried out, are not used. Yet,

preemptive patenting may be less distortive than other forms of preemption.

(c) Licenses

Another source of profits for a patentholder, besides marketing innovations himself, stems from the right to grant licenses to others. There exists a vast amount of literature on licensing, which addresses questions as: What is the optimal number of licensees? and; What is the optimal payment structure, i.e., the optimal proportion of royalties vs. fixed fees? As said above, I will not summarize the models on these issues here. For a complete survey of the literature on licensing, the reader is referred to Kamien (1992). As opposed to voluntary licensing, the requirement of compulsory licensing can be considered a restriction on the rights of a patentholder. Tandon (1982) develops a model, based on Nordhaus (1969) and Arrow (1962), where the welfare effects of compulsory licensing are studied. His results suggest that the use of compulsory licensing can be socially efficient. Additionally, he finds that, given compulsory licensing, infinitely lived patents are optimal. Tandon's model can be considered as being basic to the models on patent breadth discussed above, in the sense that it is the first to allow for competitors aside the patentholder.

2.5 Economic Choices in the Granting Procedure

This section will discuss the economic choices, both private and social, which are relevant during the granting procedure. It will not discuss what happens prior to the application. The R&D stage will be discussed in the next section since it is heavily affected by later stages. The amount of investment in R&D depends on the revenues which can be expected from it, which in turn depend on the opportunities provided by the patent system. There are two groups of factors that determine the expected returns from patent protection. First, the exploitation possibilities once a patent is granted. This group was dealt with in the above section. Second, the economic choices during the application procedure, which will be described in this section.

(a) Information Disclosure

The first choice of an inventing firm is whether or not to apply for patent protection. The obligation of information disclosure plays an important role in this choice. In the specification of the invention, the technical information must be outlined in a comprehensible manner. This specification becomes public 18 months after filing the application. The disclosure obligation causes a trade-off for the inventing firm. As the secret technical information is revealed to competitors, these competitors can absorb the new information, perhaps at some adop-

tion cost, and use it while inventing around the patent. Does the advantage of protection, even imperfect, outweigh the disadvantage of giving up secret information?

A patent race with two patentable inventions, an intermediate and a final one, is an appropriate model to study this information issue (see Grossman and Shapiro 1987, and Harris and Vickers 1987 for these races). Scotchmer and Green (1990) use such a multistage patent race to examine the relation between disclosure decisions and the novelty requirements of the patent office. First they observe a social trade-off in setting the stringency of examination. The social goal of providing an innovation incentive is best served under strong novelty requirements, which is expected to result in higher profits for the patentholder. The social goal of early information disclosure, on the other hand, is served under weak novelty requirements. Innovations become public earlier. Strategic private considerations regarding information disclosure make this trade-off rather complex. Suppose two firms race for a final patent with the possibility of intermediate patenting. The possibility of an intermediate patent is provided by weak novelty requirements and would be absent with stringent requirements. The firm that arrives at the intermediate patent first may not want to give up this advantage in the race and thus not patent, at the risk of being leapfrogged by the other and ending up with nothing at all. Another advantage of not patenting is that the other firm might drop out the race when it is signalled that there is too large a gap. Taking this type of strategic behaviour into account, Scotchmer and Green conclude that in general weak novelty requirements are socially preferred.

Another theme in their article is the rule that is applied in the case of disputes. When two firms simultaneously claim very similar inventions, which firm is to be granted the patent? There are two rules to determine the patentability. The first-to-invent rule says that only the first invention, from a historical perspective, is patentable. The first-to-file rule says that only the invention that was first reported through an application file is patentable. Scotchmer and Green find that, relative to the first-to-file rule, the first-to-invent rule discourages information disclosure. The reason being that, with first-to-invent, a first innovator tends to wait with claiming rights until other inventors show up. This way he prevents early spill-overs from disclosure while the patent can still be obtained.

In addition to the disclosure of technical information, the disclosure of private market information, only present with the inventor, on the possible profits of a competitor can play a role. Horstmann, MacDonald and Slivinski (1985) examine the signalling effects of a patent application. Their assumption is that a patent does protect against exact duplication but not against inventing around in the form of imitations. The profitability of duplication (in the absence of a patent) and imitation by a competitor is exactly known by the inventor. The competitor

only knows the distribution of possible profits. But the competitor can extract information from the inventor's patenting behaviour. Several conclusions on patenting and imitating behaviour can be drawn from their signalling model. For instance, if an imitation is not profitable, the inventor will always patent. Furthermore, if the inventor patents, the competitor will not enter the new market. Finally, if the inventor does not patent, the competitor will imitate.

It should be stressed here that patents are just one instrument to protect inventions and their associated profits. Alternative instruments are, for example, secrecy, exploiting the lead time and descending the learning curve. Secrecy implies no information disclosure. This alternative is more relevant to process innovations than to product innovations which enter the market physically. Lead time exploits the temporary information advantage. Competitors need time to study the invention and to be able to invent around.

(b) Permissive Patents

I have already pointed out that a patent is strongly exclusive and enables the holder to prohibit others to exploit similar inventions, even when generated independently. This right starts at the moment priority is obtained, which is usually on the application date. Thus, if two firms are involved in a patent race for an invention, only the firm that finishes first and applies for a patent will possibly obtain one. It may happen that two firms compete in R&D for several years with the race ending in a minor advantage for the eventual winner. The R&D expenditures of the loser cannot be recouped. A great deal of literature exists on the social costs and benefits of patent races (see Reinganum 1989 for an excellent overview). Aside from the advantage of increased chance that the invention will be generated more rapidly, duplicative R&D is always wasteful because identical information is generated twice. The dominant conclusion in these models is that an industry spends too much R&D from a social welfare point of view. The major reason is that if one firm increases its R&D, this diminishes the chance that the other firm wins the race. The first firm, however, does not take this effect on the second into account. Therefore, the common result is that the R&D expenditures at industry level (of the two firms together) are too high from a social welfare point of view (for a good example of a patent race model see Loury 1979).

La Manna, MacLeod and de Meza (1989) investigate whether a less strong form of patent exclusivity is socially beneficial. A patent system is studied where not only the winner of a race obtains a patent but also later finishers. A time limit, starting from the winner's application date onwards, could be set within which later finishers can still apply. The major trade-off occurring is that a losers-take-some reward structure diminishes the expected return from R&D because

competition in the post R&D stage will drive profits down. At the same time, however, this competition will benefit consumers. A strict regime where the winner of the race takes all thus encourages R&D but impedes diffusion, while a permissive regime with multiple prizes discourages R&D expenditures but promotes diffusion. La Manna et al. find that, under certain conditions, the permissive regime¹¹, which allows for genuine but late finishers, is preferred to the strict regime, which prevents late finishers.

(c) The Strategic Role of the Patent Office

All models discussed so far describe a scenario where the patent office first sets the length and/or scope of protection and where the potential innovators react to this design later. In a broader framework, the patent office and an innovator could be considered as players in a patent regulation game. The game most commonly described in the patent literature is one where the patent office is the Stackelberg leader, who first chooses his strategy, and the innovator is the Stackelberg follower, who reacts to the leader's choice. For example, in Klemperer (1990) the patent office first chooses a mix of patent breadth and length and an innovator is supposed to react optimally to this mix. La Manna (1992) examines the possible strategic roles of the patent office in this regulation game. Instead of Stackelberg leader, the patent office in his model can choose to be the follower. In a game where the patent office has the role of follower, it provides a set of mixtures which are welfare maximizing and the innovator chooses from this set the profit-maximizing mixture. The patent office has two instruments at its disposal: patent height and lifetime. Height is given by a minimum cost reduction or a minimum outward shift of the demand function. A traditional patent game would be organized the other way around: the innovator has several mixes of patent height and length and the patent office chooses the socially optimal mix. One of the conclusions of La Manna (1992) is that for certain cost and demand parameters, the traditional role of the patent office as leader yields lower welfare levels than its role as follower.

2.6 Feedback to the Innovation Incentive

The basic economic intention of patents is that they reduce or, preferably, close the gap between the social and private value of an invention. An exclusive right provides the patentholder with market power and thus more revenues than would be possible without a patent. This extra profit, labelled as V here, is the

11. The permissive regime could be interpreted as weak novelty requirements, or as a utility patent.

carrot which induces a potential inventor to invest in R&D. Two questions are important for the design of optimal innovation incentives: (i) how large a V must be provided for the potential inventor?; and (ii) how can V be provided most efficiently?

(a) *The Size of the Incentive*

There is the natural problem in R&D incentive design that the extra profit V differs per invention. Nordhaus (1969) already pointed at the disadvantage of uniform patent lives. He has shown that the theoretically optimal life of a patent depends, among other things, on the price elasticity of demand for the product. Products with lower elasticity of demand should have higher optimal lifetimes. On the cost side, technological opportunities affect the optimal duration. Industries where many technological opportunities are present and inventions are thus easier, should have smaller optimal patent lives. As Nordhaus (1969) also remarks, invention-specific patent protection can hardly be realized.¹² Two reasons can be given for this. First, the information needed to estimate the social and private value of an invention is costly to process. Second, even if this information could be processed, it would be hard to obtain because inventing firms probably would have a tendency to overestimate the social value of their inventions in order to get more protection and profits. Although in practice optimal patent lives are very hard to determine for individual inventions, one could say that the scope of patent protection naturally adjusts for the importance of inventions. A basic invention can reasonably be expected to yield more social value than a minor improvement. Since more claims can be formulated for the basic invention, a wider protection is obtained, resulting in higher profits for the basic inventor. The basic invention then not only has higher social value but also more protection and higher private value than the improvement.

A deviant argument with respect to incentive size is advanced in Hirshleifer (1971). His article explains the pecuniary economics of invention. Contrary to technological effects of invention, by which are meant improvements in production functions, pecuniary effects are "the wealth shifts due to the price revaluations that take place upon release and/or utilization of the information" (Hirshleifer, p. 571). The original inventor is in the unique position to forecast the pecuniary effects of his invention. A major source of profit for an inventor lies in capturing these pecuniary effects. Inventions can thus be made profitable, even *without* patent protection. By providing patent protection and the associated extra profits, on top of the pecuniary effects, an inventor could well be overcompensated.

12. The institution of utility models ('petty patents') with about 6 years of protection for smaller inventions, however, is a - primitive - form of invention-specific lifetime.

sated. To illustrate this, the future introduction of a certain product innovation can make the profits of firms producing current substitutes decrease. The product innovator could profit from this information by buying put options on the stock of his future competitors. R&D expenditures could thus be recaptured and a patent would not be necessary.¹³

(b) *The Incentive Structure*

Question (i) in the introduction of this section on the optimal size of V can be avoided by reformulating question (ii) as: how can the welfare loss of providing one unit of V be minimized? Kaplow (1984) introduces, in a mainly legal article, a ratio test as an instrument to judge monopolistic practices of patentholders. He proposes the ratio of the patentee's reward divided by the monopoly loss as a guideline for patent and competition policy. In general, he says, the higher this ratio, the more desirable, from a social point of view, the monopoly practice is. After formalization of the problem for the patent office to set an optimal mixture of breadth and length, Klemperer (1990) comes up with an identical (inverse) ratio of welfare loss divided by patentholder's profit as the objective function to be minimized. Thinking in terms of this ratio proves to be very useful in patent questions. The difficult question of incentive size is avoided and the focus can be placed on the structure of the incentive. As explained in section 2.4, both Klemperer (1990) and Gilbert and Shapiro (1990) assume an exogenous incentive size and search for the conditions under which the ratio of welfare loss to profit is minimized.

The same procedure is followed in chapter 4 of this book in order to examine the practice of price discrimination by a patentholder. There it is shown in a model of patent breadth that, given linear demand functions, the innovation incentive in the form of a fixed profit can be provided more efficiently by allowing for price discrimination by the patentholder. In the price discrimination literature (see Varian 1989 for an overview), it is a well-known result that price discrimination, as compared to uniform pricing, lowers static welfare for linear demand functions (given that no new markets are served under price discrimination). Because profits always increase under price discrimination, it is not clear beforehand whether the ratio of welfare to profit is larger under price discrimina-

13. An example of this kind was the news that Xerox developed a new colour film, in which silver is no longer used (*Volkskrant*, November 24, 1993). The photography industry is the largest buyer of silver and represents about 30% of total demand. As a result of this news, the price for silver on the international market dropped by 5% within two days. Xerox could have foreseen the impact of its new film on the silver price. If Xerox had managed to buy put options on silver before the news became public or other parties noticed this signal, it could have realized some rents.

tion or under uniform pricing. The analysis in chapter 4 shows that the ratio under discrimination is equal to the ratio under uniform pricing if there is no explicit model of patent protection underlying the demand functions of the patentholder. From the perspective of providing an innovation incentive, the practice of price discrimination is thus neutral to uniform pricing. If an explicit model underlies the extent of patent protection and the resulting demand in each market, then the practice of price discrimination is shown to be a socially more efficient manner to provide an innovation incentive.

(c) The Patent as Incentive

As already pointed out above, the literature on the topic of patent races is extensive.¹⁴ A patent race describes the technological competition between two or more firms aiming for the same innovation. The patent race is often modelled as a winner-takes-all game (Reinganum 1989). The winner of the race obtains a patent and the associated monopoly profit while the loser gets nothing at all. The solution of such games is often characterized by an overinvestment, from a social point of view, in R&D at industry level. As explained in section 2.5, the major reason for this result is the fact that one firm in the race does not take into account the negative effect on other firms if it increases R&D compared to the others. This is labelled in the literature as the 'common pool problem'. Mortensen (1982) shows that the common pool problem in patent races can be corrected. If the winner of a race were required to compensate each of the losers for the lost values of prospect, then the equilibrium in the patent race is socially efficient. The strong exclusivity of patents can be said to cause the industrial overinvestment. Modification of the exclusive character corrects this tendency.

Wright (1983) compares the performance of three instruments that provide innovation incentives: patents, prizes and research contracts. A patent is awarded to one successful inventor, a prize to all successful inventors. A research contract purchases units of research activity beforehand at a fixed price. As shown above, patents provide the inventor rewards at the cost of temporary deadweight losses. Instead of indirectly via patents, an equivalent monetary compensation could also be offered directly to researchers. This compensation could be funded by less distortionary means, compared to deadweight loss causing patent monopolies. Put this way, there is no reason to assume that patents are superior to prizes or research contracts (the monetary compensations) in order to induce invention. Wright points out that this conclusion on the disadvantage of patents may

14. This patent race literature has developed simultaneously with the auction literature in game theory. Patent races are generally considered to be fine applications of auction theory. This may explain the extensive literature on patent races.

depend on an assumption of symmetric information, both present with an administrator and researchers, concerning the costs and benefits of research. In a context of asymmetric information on the cost and probability-of-success functions, any of the three alternatives may be the best choice. In brief, some of the considerations of Wright are the following. Compared to research contracts, the disadvantage of both prizes and patents is that they induce the common pool problem (socially excessive resources to innovation). Where this can be corrected in the case of symmetric information (also found in Mortensen 1982), it cannot be corrected for asymmetric information scenarios. However, the advantage of prizes and patents is that they are more decentralized invention incentives that use private information exclusively available to researchers. In this respect, patents perform even better than prizes because the information on the benefits of success is also incorporated, whereas the benefits of prizes are fixed. The disadvantage of patents, however, is the incurred deadweight loss.

2.7 Summary and Concluding Remarks

This chapter has provided an overview of the models on patents that have appeared in theoretical Industrial Organization. The purpose of this chapter has been to estimate the state of the art and to see where niches can be filled. As a start, I have explained the basic economic arguments of patents and other intellectual property. Next, I have briefly outlined the patent-granting procedure of the European Patent Office. This procedure has formed the organizing principle behind the survey. More precisely, the theories were discussed in the reverse chronological order of a typical patent procedure.

First, the stage of exploitation of a granted patent was examined. Models on the length, breadth and height of patent rights were discussed, as well as models dealing with the restrictions, particularly stemming from competition policy, that a patentee faces when exploiting his rights. Next, models relevant to the procedural stage were reviewed. The most important decision an applicant has to make in this stage is whether or not to reveal secret information to his competitors. Another topic discussed is the normative question whether or not to allow later applicants after the first applicant. Finally, the role of the patent office in the patent regulation game was elaborated. The final stage dealt with was the stage of competition in R&D: How does the prospect of obtaining a patent affect the stage of R&D? Table 2.2 provides an overview of the topics and articles that are discussed in this chapter.

Table 2.2

Overview of the theoretical patent literature

TECHNOLOGICAL COMPETITION	
Innovation Incentive	Reinganum (1989): patent races Mortensen (1982): efficiency Kaplow (1984): incentive structure
Alternatives	Hirshleifer (1971): information Wright (1983): prizes, subsidies
GRANTING PROCEDURE	
Information Disclosure	Horstmann et al. (1985): signalling Scotchmer and Green (1990): novelty
Permissive Patents	La Manna et al. (1989): application date
Role Patent Office	La Manna (1992): rules
EXPLOITATION	
Rights: Length	Nordhaus (1969): R&D monopoly Kamien and Schwartz (1974); De Brock (1985): R&D rivalry Goyal and De Laat (1991): multiple inventions David and Olsen (1992): learning economies
Breadth	Klemperer (1990); Waterson (1990); Gallini (1992): one market Kitch (1977); Merges and Nelson (1992); Matutes, Regibeau and Rockett (1993): more markets Chapter 4: price discrimination
Height	Scotchmer and Green (1990): disclosure Chapter 3: improvement Chapter 5: breadth
Restrictions	Gilbert and Newberry (1982): preemption Hausman and MacKie-Mason (1988): price discrimination Gilbert and Shapiro (1990): maximum price

Most of the models that were discussed and that will be advanced in the following chapters use a static, partial equilibrium framework. Partial equilibrium analysis allows for more and richer detail. The same might be said about static analysis; since the proceeding of time is not incorporated, more focus can be placed on the static moment. Although many useful insights emerge from these static and single market models, the inclusion of time into the analysis could prove to be very illuminating, especially in the context of technical progress. Some recent analyses use dynamic, general equilibrium models. An early analysis is Judd (1985) on product innovation. But this article had its major impact outside the patent literature; it can be considered as one of the first new growth models. Other analyses are Chou and Shy (1993), on the length of patent protection and its impact on new product development afterwards, and Helpman (1993), on the issue of global patent protection in the Uruguay Round. Future research can examine many known patent questions in a dynamic and general equilibrium framework. But it is felt that static and partial analyses continue to be useful, particularly when looking for and examining new questions.

Table 2.2 also gives an impression of the topics that will be discussed in the following chapters of this book. As mentioned in section 2.5, patent height has received surprisingly little attention in the literature, despite its revealed importance in industry and patent practice. Chapters 3 and 5 provide the first steps to fill the gap. Chapter 3 will analyse patent height as an isolated dimension, while chapter 5 will study height in combination with the breadth dimension. The intersection of the breadth dimension with competition policy, more precisely the policy regarding price discrimination, is examined in chapter 4.

Patent Height and Competition in Product Improvements

Empirical studies on the appropriability of inventions (for example, Mansfield et al. 1981, Levin et al. 1987 and Griliches 1990) provide evidence that patent protection is limited. As mentioned in chapter 1 (fact b), many opportunities are left open for competitors to invent around patents. One can therefore hardly claim, as is commonly done, that patents provide pure monopoly power. They restrict at most the opportunities for inventing around. The legal profession is far more aware of this fact. Lawyers are often confronted with the imperfectness of patent protection as the considerable amount of case law from patent disputes demonstrates (see Cornish 1989 as a leading textbook on intellectual property).

Until recently, theoretical economists have hardly taken into account the economic and legal practice of patent protection. Nordhaus' (1969) seminal work on patent length, for example, assumes perfect protection during the lifetime of the patent. The same assumption is also implicitly present in most models of technological competition where the first firm that generates an invention obtains a patent and takes all the profits, leaving the loser of the race no chances whatsoever (for an overview of patent races, see Reinganum 1989). Recent contributions aim at relaxing the assumption of perfect protection (Gilbert and Shapiro 1990, Klemperer 1990). Both studies allow for imitations beside the patent, provided that these imitations are not too similar. When substitutes in the form of imitations are less similar to the patented product, the monopoly power of a patentholder increases. Gilbert and Shapiro are therefore able to define the breadth of patent protection rather generally as the ability of the patentholder to raise price. Their focus is on the socially optimal mixture between patent length and breadth. As extensively discussed in chapter 2, Klemperer defines patent breadth more explicitly within a model of horizontal product differentiation, as an interval where no competitors are allowed. The usage of a model of horizontal differentiation turns out to yield richer conclusions about the optimal mixture, reproducing and extending those of Gilbert and Shapiro.

In this chapter, I focus on the height of patent protection. As outlined in chapter 2, patent height is mainly determined by the stringency of the novelty requirements used by examiners in the patent office. Most patent laws dictate that an invention is only patentable if it is non-obvious and involves an inventive step. These conditions are certainly satisfied for the small group of fundamental inventions. But the judgement on patentability is more difficult where improvements of existing products are concerned. When is an improvement obvious and how much inventivity is required? It is the task of patent examiners to define the bounds of patentability and infringement of current patents. The central question in this chapter is how the height of patent protection affects the competition in product improvements in a duopoly where one firm holds a patent, which restricts the strategy set of the other.

Before stressing why the distinction between patent height and breadth is important, let me first explain the difference more exactly. Whereas breadth is defined as the extent of protection against imitations, height indicates the protection against improvements. In terms of product characteristics, one could say that height defines the minimal number of new elements which an invention must contain, whereas breadth defines the maximum number of elements which are allowed to be imitated. Suppose a patented product is made up of v elements. Think, for example, of such technical attributes as materials, design, electronics schemes, etc., which define the invention in a unique way. Protection height then indicates how many new elements h are required in order not to infringe an ongoing patent. Breadth defines the maximum number b out of the v that is allowed to be imitated without infringing.

The difference between imitations and improvements can be found on both the cost and the demand sides. Although imitations may also require R&D investment in order to obtain learning and imitation capabilities (Cohen and Levinthal 1989), improvements are likely to cost more because new information has to be generated. With regard to the demand side, as previously explained, Klemperer (1990) treats imitations as horizontal differentiations of the patented product. Consumers do not agree on the most preferred variety. Some consumers prefer protected varieties, while others prefer imitations outside the protected interval. Improvements are basically different from imitations in this respect. Improvements can be considered as vertical differentiations. It is reasonable to expect that, at equal prices, all consumers prefer an improvement to an older version of the product. Patent height can then be thought of as defining a protected region on the vertical product spectrum. Competitors that enter the market are restricted in their choices of improvement. They must generate a

minimum level of improvement in order not to infringe the current patent.¹ The dimensions of breadth and height thus define protection against two different types of competition; one resulting from horizontal differentiation and another resulting from vertical differentiation. These two types of competition have a differential impact on the patentholder. The effect of vertical differentiation is that it relaxes price competition; therefore firms have a tendency to differentiate their products maximally (Shaked and Sutton 1982). Horizontal differentiation may have an opposite effect. At least for linear transport costs, as in the original Hotelling (1929) analysis, firms differentiate minimally.

Taking these cost and demand features into account, two basic strategies for inventing around a patent can be distinguished: an imitation strategy 'aside' the patented product and an improvement strategy 'above' it. An example may clarify the two possible strategies. Klemperer (1990, p. 114) mentions the patent of Prince Manufacturing on the oversized tennis racket. The breadth of protection is known to run from 85 to 130 square-inches rackets. A competitor pursuing an imitation strategy can only enter with a similar racket outside this protected range. But there is also another possibility, which is not examined by Klemperer. A competitor can improve a racket inside the protected range in such a way that it satisfies patent examiners. For example, a 100 square-inches racket can be made stiffer by the use of a new fibre or a better design. Both cost and demand properties can play an important role in choosing the appropriate inventing-around strategy. An imitation strategy has a relative cost advantage but may be more severely restricted by patent breadth because, being a horizontal differentiation, the imitation may be chosen close to the patented product. A competitor that pursues an improvement strategy has a natural tendency to create distance away from the patent, but this tendency may be suppressed by high R&D costs. The impact of the imitation strategy on the patentholder's profits has been examined in the above-mentioned models of patent breadth. So far, however, the impact of the improvement strategy has not been analyzed.

This chapter will concentrate on the effects of novelty requirements once a patent is granted. Scotchmer and Green (1990) have analyzed a related problem. They discuss the social trade-off that occurs in setting the novelty requirements when firms apply for patents. Novelty requirements appear to have a dual character. They determine the coverage of granted patents - the study object of this chapter -, but they also affect the stage of R&D competition before the patent is

1. The literature on minimum quality standards also uses vertical differentiation models (e.g., Ronnen 1991 and Crampes and Hollander 1991). The major difference with the analysis presented here is that only the non-patentholder is restricted in my model whereas in the other models mentioned both firms are restricted.

possibly granted. This second aspect is studied in Scotchmer and Green (1990). The main benefit of weak novelty requirements then is that technical information becomes public at an earlier date so that other firms can build further on it sooner. However, weak requirements may also induce the inventor to withhold his invention, thereby keeping the lead, and aim at a more profitable invention later on. Strong novelty requirements reduce the amount of wasteful duplicative R&D by inducing firms that have fallen behind to drop out of the race. In order to focus fully on the extent of protection, I will omit the stage of R&D competition here.

The plan of this chapter is as follows. In section 3.1, a model is constructed which is based on the idea of improvements being vertical differentiations of a new product. First, I examine the competition in product improvements in the absence of patents. Section 3.2 will analyze the effects of patent height on the competition in product improvements. In a nutshell, the major conclusions are that low patent heights do not affect the natural market equilibrium without patents, medium heights can be favourable and unfavourable for the patent-holder, and high patents provide pure monopoly power. A short welfare analysis is carried out in section 3.3. Section 3.4, finally, summarizes and provides some directions for future research.

3.1 The Model

The starting point of the analysis is the appearance of a basic invention. In this section, I assume that this basic invention is public property. The basic invention is so raw that no consumer is prepared to buy it. Consumers can only extract surplus from the basic invention if it is further developed. Defining the demand side more precisely, the net surplus U of an individual consumer is given by the following (indirect) utility function:

$$U = \begin{cases} mv - p & \text{if the consumer buys} \\ 0 & \text{otherwise} \end{cases} \quad (3.1)$$

where $v \in [0, \infty)$ is the degree of improvement (with the basic invention given by $v = 0$); p is the price of improvement v and m is a parameter indicating the intensity with which a consumer prefers improvements. The parameter m is uniformly distributed with density 1 on $[z - 1, z]$, where $1 < z < 2$. A consumer only buys if he gets a non-negative net surplus. Each consumer decides in each period (the model will encompass more periods) whether or not to buy (one unit). The utility function (3.1) is standard in the literature on vertical differentiation (see Shaked and Sutton 1982). Applying this function in the context of technical

change is justified because consumers are expected to evaluate product improvements in much the same way as qualities of a product. Consumers prefer large to small improvements, just like they prefer high-quality to low-quality products.

Two firms, 1 and 2, are potentially entering the new market with improvements of the basic invention. Each of the firms is assumed to offer one improvement at the most. There are two motivations behind this assumption. First, the diversity of consumers' tastes (given by the range of m) is chosen such that there is only room at the market for two improvements. Competition analysis then requires each of the two firms to have one improvement at the most. Second, it is true that in business practice one can often observe that a firm does not have one patent but a number of related patents, all in the same technical field. But even such a cluster of patents cannot provide perfect protection. Competitors can still invent around the cluster. Essentially we are then back at the scenario studied in this chapter: one firm has a patent (or a cluster of patents) and another firm enters the market with an improvement (of the most advanced patent of the cluster). Once chosen, I assume that an improvement cannot be changed. As we will see below, there are circumstances under which it would be beneficial for a firm to lower its improvement level. However, high fixed costs associated with each improvement are assumed to prevent this (these costs are made before the start and do not enter the analysis below).

The competition between both firms takes place in two stages. In the first stage, each firm decides whether or not to enter the market and which improvement to choose. Knowing which decisions are made in the first stage, each firm then chooses a price strategy. This sequence reflects the fact that price decisions are more flexible than improvement and entry decisions. The solution concept that I employ for this two-stage model is the subgame perfect equilibrium. It can be obtained by backward induction. First, I will determine the Nash equilibrium in pricing strategies of the second stage. In the second stage, I will determine the equilibrium strategies in product improvements and entry. Only pure strategies are considered.

It is assumed here that both firms are characterized by the same R&D cost structure, which is kept as simple as possible. There are no development costs involved in the generation of improvements. It only costs time. Starting at time 0, firm i introduces v_i in period T_i , where $T_i = \alpha v_i$, for $i \in \{1, 2\}$ and $0 \leq \alpha \leq 1$. The profit that is lost because of a later introduction time is the real development cost. If both firms enter the market, the first-stage subgame is defined by a pair of improvements $\{v_S, v_L\}$, such that $v_S \leq v_L$. The subscript S denotes the firm that has developed an improvement that is smaller than the other improvement. The other firm has the larger improvement and is denoted by subscript L . Note that each firm can develop either the small improvement or the large improvement.

Assuming that both firms start their development in the same period, v_L is always introduced later than v_S . As a consequence, the market structure can change over time. The firm with v_S will enter in period T_S . It then has a temporary monopoly position and sets a monopoly price, denoted by p_M . This monopoly will last until the other firm enters at T_L with v_L . From T_L until a doomsday J the market structure then will be a duopoly. By assumption, the former monopolist can only react to the entrant by setting a new price, denoted by p_S , as a function of the entrant's price p_L . Assuming that all consumers are price and improvement takers, during monopoly each consumer has the choice between buying v_S at price p_M or not buying at all. Consumer choices extend when the later firm introduces v_L at T_L . Each consumer can then choose to buy v_S at price p_S , v_L at price p_L , or none of the improvements.

Price Competition

The following demand function is relevant in two cases. First, in the scenario in which only one firm enters the market and remains a monopolist until doomsday J . Second, during the period when the small improvement firm is alone in the market and the other firm has not entered yet.

$$x(p_S) = z - p_S/v_S \quad (3.2)$$

for $p_S/v_S > z - 1$. Without affecting the basic results, I assume that the unit production costs are zero. It can be shown that the monopoly price and profit per period are $p_M^* = zv_S/2$ and $\pi_S(p_M^*) = z^2v_S/4$. The lower market segment $[z - 1, z/2]$ is not served.

Now consider the demand functions in case of duopoly. The consumer who is indifferent during duopoly as to the small and large improvement is given by the preference parameter $\mu = (p_L - p_S)/(v_L - v_S)$. The firm with v_S , which charged p_M before, now changes its price and charges p_S . For computational ease, I assume that the market is served completely during duopoly and take $(v_S + 2v_L)/(2v_S + v_L) < z < 2$.² The demand functions are then:

$$x_S(p_S) = \mu + 1 - z, \quad x_L(p_L) = z - \mu \quad (3.3)$$

The Nash equilibrium prices are $p_S^N = (2 - z)(v_L - v_S)/3$ and $p_L^N = (z + 1)(v_L - v_S)/3$, with associated profits of $\pi_S(p_S^N) = (2 - z)^2(v_L - v_S)/9$ and $\pi_L(p_L^N) = (z + 1)^2(v_L - v_S)/9$ (see Tirole 1988 for an elaboration).

2. This condition turns out to be always possible in the analysis to follow. All equilibrium values of v_S and v_L fulfil the condition that $1 < (v_S + 2v_L)/(2v_S + v_L) < 2$.

Competition in Product Improvements

First, I will examine the pure monopoly scenario, where the monopoly position lasts from the introduction time until the doomsday J . The monopolist will have profits $\pi_s(p_M^*)$, as determined above. The profit function of such a monopolist in the first stage is:

$$\pi_s(v_s) = (J - T_s)\pi_s(p_M^*) \quad (3.4)$$

In order to keep the analysis as simple as possible, I assume that pay-offs are not discounted. A similar effect originates from the doomsday J , when everything ends (for example because a complete new product is generated that takes all demand). The monopolist is confronted with the following trade-off: his profit per period increases if he develops a larger improvement, but this costs more development time so that there will be less time to capitalize on his investments. The next step is to calculate the optimal improvements as a function of each other, given the equilibrium prices determined above. After substitution of the time cost structure $T_s = \alpha v_s$ and the optimal profit $\pi_s(p_M^*)$ in (3.4), the optimal improvement of the monopolist can simply be determined as $v_s^* = J/(2\alpha)$, with the associated optimal profits of $\pi_s(v_s^*) = z^2 J^2 / (16\alpha)$. The optimal improvement v_s^* increases if the monopolist is given more time (later J) and if the development costs are lower (smaller α). The optimal monopoly profit $\pi_s(v_s^*)$ also increases if the demand properties are more favourable; a higher z means that improvements are more valued by consumers.

When two firms enter the new market, the profits of the firm with v_s contain monopoly profits, gained from T_s until T_L , and duopoly profits, from T_L until doomsday J :

$$\pi_s(v_s; v_L) = (T_L - T_s)\pi_s(p_M^*) + (J - T_L)\pi_s(p_s^N) \quad (3.5)$$

The profit function of the firm that enters later at T_L with v_L is:

$$\pi_L(v_s; v_L) = (J - T_L)\pi_L(p_L^N) \quad (3.6)$$

Before proceeding, I want to emphasize some important properties of the profit functions (3.5) and (3.6). The profit per period $\pi_s(p_M^*)$ in equation (3.5) increases in v_s . This property was already mentioned above in the pure monopoly case. In order to simplify the notation, let $q \equiv z^2/4$, so that $\pi_s(p_M^*) = qv_s$. The duopoly profits per period $\pi_s(p_s^N)$ and $\pi_L(p_L^N)$ decrease in v_s for the small improvement firm (check 3.5) as well as the large improvement firm (check 3.6). This property is well known in the vertical product differentiation literature and is the result of

more relaxed price competition as products are more differentiated (Shaked and Sutton 1982). It can also be found in the way profits behave in v_L . Since a larger v_L means a larger distance between both improvements, the profits per period increase for both firms. Define $r \equiv (2 - z)^2/9$ and $s \equiv (z + 1)^2/9$, so that $\pi_s(p_s^N) = r(v_L - v_s)$ and $\pi_L(p_L^N) = s(v_L - v_s)$. For the values of z taken here it holds that $s > q > r$. This ranking will turn out to be important further on.

After substitution of the introduction times and the profits in the second stage in the profit functions (3.5) and (3.6), the following improvement reaction functions can be derived:

$$v_s^*(v_L) = v_L(q + r)/(2q) - rJ/(2\alpha q) \quad (3.7a)$$

$$v_L^*(v_s) = v_s/2 + J/(2\alpha) \quad (3.7b)$$

where $v_s, v_L \in [0, J/\alpha]$. An important property of (3.7a) is that it implies that π_s increases monotonically in v_s along the reaction curve. This can be checked by substituting (3.7a) into (3.5) and taking the first derivative of π_s with respect to v_s^* , which turns out to be positive. The intuition behind this is the following. If v_L increases, v_s^* also increases, but by less (since $(q + r)/(2q) < 1$). The number of monopoly periods as well as the monopoly profit per period increase. The profit per period in duopoly also increases because the distance between v_L and v_s becomes larger. These positive effects on π_s outweigh the only negative effect of a smaller number of periods before doomsday. The opposite holds for the firm with the large improvement; π_L decreases monotonically in v_L along the reaction curve (3.7b). The intuition is simple. Suppose v_s increases. As a consequence, v_L^* will increase but only by half as much, so that the distance between both improvements becomes smaller. Considering, in addition, that the remaining time before doomsday shrinks, it is clear that π_L decreases.

The Nash equilibrium in the first stage of competition in product improvements can be found by solving the reaction functions (3.7a) and (3.7b) simultaneously for v_s and v_L . Define this solution as the *basic* Nash equilibrium $\{v_s^N, v_L^N\}$, given by:

$$\begin{aligned} v_s^N &= J(q - r)/(\alpha(3q - r)) \\ v_L^N &= J(2q - r)/(\alpha(3q - r)) \end{aligned} \quad (3.8)$$

Between T_s and T_L , the upper segment of the market $[b/2, b]$ is served by the temporary monopolist with improvement v_s^N at a price of $p_M^* = zJ(q - r)/(2\alpha(3q - r))$. When the other firm enters in period T_L with the large improvement v_L^N , it

serves more customers than the monopolist. The upper market segment, given by $[(2z - 1)/3, z]$, is served at a higher price of $p_L^N = ((2z - a)Jq)/(3\alpha(3q - r))$. The small improvement firm now serves the lower market segment $[z - 1, (2z - 1)/3]$, which was not served before, at a price of $p_S^N = ((2z - 1)Jq)/(3\alpha(3q - r))$.

In this basic Nash equilibrium, the overall profit of the firm with the small improvement ($\pi_S(v_S^N)$) turns out to be positive and lower than that of the firm with the large improvement ($\pi_L(v_L^N)$): $q^3J^2/(\alpha(3q - r)^2) < sq^2J^2/(\alpha(3q - r)^2)$, since $s > q$. It is not clear in advance which firm will have the large improvement. There are in fact two basic Nash equilibria in pure strategies. One in which firm 1 has the small improvement and firm 2 has the large improvement $\{v_1 = v_S^N, v_2 = v_L^N\}$ and one where this division of roles is reversed $\{v_1 = v_L^N, v_2 = v_S^N\}$.

3.2 The Effects of Patent Height

This section studies scenarios where the basic invention is privately owned. One of the two firms, say firm 1, holds a patent on the basic invention. Two cases are studied: [1] firm 1 cannot keep the basic invention secret, not even for a very short period of time; and [2] firm 1 can keep the basic invention perfectly secret. Any level of the first improvement can be protected by a patent. The patent system under consideration is assumed to be a 'first-to-file' patent system. The patent protection is of infinite length and of height h .³ It is assumed that the criterion for judging infringement is precise and known by the patentholder and his competitor.⁴ The meaning and effects of the patent height h are explained below. To simplify the notation without affecting the results, I assume that $J = 1$ and that $\alpha = 1$.

Secrecy is Impossible; Patent on the Basic Invention

In scenario [1], firm 1 is assumed to patent the basic invention immediately (given the lack of secrecy and the first-to-file rule, firm 2 may do so otherwise). I will come back to this assumption below. The technical information in the patent

3. Since competitors are assumed to choose an improvement of and not an imitation of the patented new product, the breadth of protection is not important here. Case law seems to indicate that the novelty requirements overrule the allowed similarity of imitations (Cornish 1989). In other words, an identical variety is permitted provided that it is improved sufficiently. Mixtures of patent breadth and height as well as optimal strategies for inventing around are studied in chapter 5.

4. Waterson (1990) examines a patent system where it is not clear beforehand whether the patentholder or the possible infringer wins in court. This uncertainty may affect the patenting decision.

application becomes public and causes both potential entrants to start their development work with the same knowledge level. Hence, there is no asymmetry as far as knowledge is concerned. However, there is asymmetry because firm 1 has a patent on the basic invention and firm 2 has not. The patent is not a barrier to entry that cannot be overcome. It only protects a limited area beyond the basic invention. More precisely, the protection of the patent with height h is the region $[0, h]$.⁵ The non-patentholder is not allowed to choose his improvement in this region. He has to choose his improvement in the free, non-protected region $[h, 1]$.

The second stage of price competition is not affected by the institution of patent heights. The first stage of competition in product improvements, however, can yield different results. Both firms continue to choose their improvements simultaneously. The strategy space of the non-patentholder 2 shrinks to $v_2 \in [h, 1]$. The strategy space of firm 1 remains unchanged $v_1 \in [0, 1]$. This asymmetric effect of patent height is the major difference with the models on minimum quality standards of Ronnen (1991) and Crampes and Hollander (1991). In their models, the minimum standard symmetrically limits the strategy spaces of both firms. With firm 1 holding a patent, each of the firms can still choose the small or large improvement in equilibrium. But we will see that, depending on the height of protection, the improvements can become larger compared to those in the basic Nash equilibria. We will also see that in one category of heights, one basic Nash equilibrium will not occur.

The effects of patent height can be divided into five categories. The two categories for extreme values of h are obvious. Low patents ($0 \leq h \leq v_s^N$) do not affect the basic Nash equilibria. Firm 2 can still have the small improvement without being restricted effectively. In the other extreme category of high patents ($h \geq 1$), firm 2 does not have enough time to develop an improvement that does not infringe the patent of firm 1. The patentholder is then assured to have a monopoly position until doomsday 1 and the only improvement that will be on the market is $v_1 = 1/2$. For the medium values $v_s^N < h < 1$, three categories can be distinguished (see Propositions 3.1.a and 3.1.b). Two categories contain heights for which firm 2 is effectively restricted in the choice of the small improvement (Proposition 3.1.a). These heights turn out to be unfavourable for the patentholder if he has the large improvement, because firm 1 loses profit compared to the basic Nash equilibrium. For firm 2, the non-patentholder, these patent heights

5. In their judgement on non-obviousness and inventiveness, patent examiners mainly rely on files of former patent applications. The body of knowledge in these files represents to a large extent the (accessible) state of the art, which serves as a benchmark for examination. This is the reason why the protection starts from 0, i.e., the patented basic invention.

are initially favourable. Proposition 3.1.a summarizes:

Proposition 3.1.a. For $h \in \langle v_s^N, \hat{h} \rangle$ there are two Nash equilibria in pure strategies (\hat{h} will be defined below): a basic Nash equilibrium with $v_1 = v_s^N$ and $v_2 = v_L^N$ and a new Nash equilibrium with $v_2 = h$ en $v_1 = v_L^*(h)$. In the new Nash equilibrium (a) both improvements are larger than in the comparable basic Nash equilibrium; (b) the profit of the patentholder 1 decreases in h ; and (c) the profit of the non-patentholder initially increases and later decreases in h . For $h \in [\hat{h}, v_L^N]$, the only basic Nash equilibrium is $v_1 = v_s^N$ en $v_2 = v_L^N$.

Firm 2 is effectively restricted in the choice of the small improvement and not in the choice of the large improvement for the heights $v_s^N < h \leq v_L^N$. What are the consequences? The one basic Nash equilibrium in which firm 1 has the small and firm 2 has the large improvement remains unchanged. The other basic Nash equilibrium, however, is no longer possible because firm 2 is not allowed to enter the market with the improvement v_s^N . The smallest improvement permitted to firm 2 is exactly equal to the height $v_2 = h$.⁶ Firm 1 will react according to (3.7b) with an improvement $v_1 = v_L^*(h)$. Note that both improvements h and $v_L^*(h)$ are larger than in the relevant basic Nash equilibrium with $v_2 = v_s^N$ and $v_1 = v_L^N$; $h > v_s^N$ holds by definition, and $v_L^*(h) > v_L^N$ because $dv_L^*/dv_s > 0$.

The optimal profit of the large improvement firm decreases in v_L along its reaction curve, as shown in section 3.2. Hence, if firm 1 has the large improvement, its profit π_L decreases in h for $h \in \langle v_s^N, v_L^N \rangle$. This can be seen in figure 3.1, where π_1' represents a lower profit level than π_1 . Since the firm with the large improvement has the higher profits in basic Nash equilibrium and these profits now decrease in h for the patentholder, there must be a value for h where the patentholder collects the same profits with the large improvement as with the small improvement, with firm 2 choosing $v_s = h$. Label this critical value for firm 1 as \hat{h} , defined as $\pi_L(\hat{h}, v_L^*(\hat{h})) = \pi_s(v_s^N, v_L^N)$. For $h > \hat{h}$, the patentholder prefers to have the small improvement instead of the large one.

What happens to the profits of firm 2 when it has the small improvement? For $h < \hat{h}$, firm 2 has to deviate from its reaction function (3.7a) and choose the small improvement larger than in the basic Nash equilibrium. As a result, its profits will initially increase in h . Profits will increase up to height h' , where h' is equal to the product improvement that firm 2 would choose if it were Stackelberg leader.

6. I restrict the analysis to *one* patent. I assumed that an improver cannot obtain a patent. Allowing for a second patent complicates matters since the basic inventor has to take into account that his strategy space might also be restricted if the improver has a patent.

In figure 3.1, π_2' represents the (Stackelberg) profit level of firm 2 given patent height h' . This profit level is higher than the one in the basic Nash equilibrium, π_2 . Profits start decreasing for firm 2 from h' on.⁷ For $h > \hat{h}$, the patentholder does not even want the large improvement anymore and leaves it to the non-patentholder, despite the fact that it yielded more profits in the basic Nash equilibrium. The economic intuition behind this surprising result is that patent height forms a credible commitment which the non-patentholder can exploit. Initially the non-patentholder is credibly committed to choosing a larger v_s than in basic Nash equilibrium. For even higher patents ($h > \hat{h}$), the commitment becomes so strong that the non-patentholder always has the more profitable large improvement.⁸

7. I limit the analysis here to the case where $\hat{h} > h'$. The maximum profits of firm 2 in the turning point do not exceed the profits with the large improvement in the basic Nash equilibrium. In the case of $\hat{h} < h'$, firm 2's profit only increases in h for $h \in (v_s^N, \hat{h})$. If the maximum profit in the turning point h' is larger than the profit associated with the large improvement in the basic Nash equilibrium, then an extra category can be distinguished in which firm 2 prefers the small improvement h and firm 1 prefers the large improvement $v_L^*(h)$.

8. This disadvantage for the patentholder may depend on two assumptions made above. First, the assumption that firm 1 patents immediately may be important here because, without the patent, the possibility would stay open for firm 1 to have the more profitable large improvement. As we will see below, however, from a critical height on, it is beneficial again for firm 1 to have the patent. By waiting and patenting later, the protection associated with an improvement patent becomes larger ($v_L \geq v_s + h$). But firm 2 could then obtain the patent as well. A critical height must exist for which both firms are indifferent to having the patent or not. It could well be the case that the profits associated with this critical height, are lower than the profits of the firm with the small improvement in the basic Nash equilibrium (e.g., if the profits associated with the restricted large improvement are lower than those associated with the patented small improvement and there is some probability of not obtaining the patent). For these cases, firm 1 would patent immediately anyway.

Another way out of the disadvantageous patent position might be licensing. Although licensing is excluded from the present analysis, I will comment on it briefly here. Note that the patentholder has an incentive to offer, even freely, the improver a license, allowing him to choose an improvement inside the protected region. Profits of the patentholder would increase. The profits of the improver and license taker would, however, decrease. Thus, unless the patentholder is able to precommit to license its patent freely, a license contract would not emerge. And even if such precommitment is possible (e.g., if the patent office recognizes the one-sided free gift and adjusts its standards of examination, or if licensing is compulsory), the improver may be able to precommit not to accept the free gift (e.g., by provoking a court or patent office judgement). If the patentholder can precommit to license freely and the improver cannot precommit not to accept, the basic Nash equilibrium would reoccur.

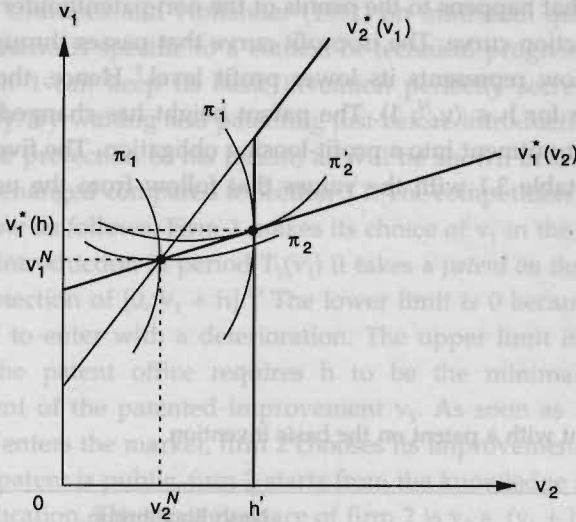


Figure 3.1 Patent on the basic invention; firm 2 has the small improvement

This advantage for the improver, however, disappears. From a certain height on, the commitment becomes an obligation; firm 2 must choose a larger improvement than it would like to. Higher patents then yield more profit for the patentholder and less for the non-patentholder if the latter is also effectively restricted in the choice of the large improvement. The following proposition holds:

Proposition 3.1.b. For $h \in \langle v_L^N, 1 \rangle$ there is one Nash equilibrium with $v_1 = v_s^*(h)$ and $v_2 = h$. The Nash equilibrium is characterized by: (a) larger improvements than in the relevant basic Nash equilibrium; (b) an increasing profit of the patentholder in h ; and (c) a decreasing profit of the non-patentholder in h .

Firm 2 is effectively restricted in its choice of the large product improvement for the patent heights $h \in \langle v_L^N, 1 \rangle$. The choice of v_L^N is not allowed and firm 2 does best choosing h . The best reaction of the patentholder is $v_s = v_s^*(h)$. The only Nash equilibrium is characterized by a small improvement of the patentholder of $v_1 = v_s^*(h)$ and a large improvement of firm 2 of $v_2 = h$. Note that both improvements are larger than in the relevant basic Nash equilibrium; $h > v_L^N$ holds by definition, and $v_s^*(h) > v_s^N$ because $dv_s^*/dv_L > 0$. In section 3.1, I have demon-

strated that the optimal profits of the small improvement firm always increase in v_s along the reaction curve (3.7b). The profits of the patentholder thus increase in h for $h \in \langle v_L^N, 1 \rangle$. What happens to the profits of the non-patentholder? He has to deviate from his reaction curve. The isoprofit curve that passes through the new Nash equilibrium now represents its lower profit level.⁹ Hence, the profits of firm 2 decrease in h for $h \in \langle v_L^N, 1 \rangle$. The patent height has changed for firm 2 from a profitable commitment into a profit-loosing obligation. The five categories are summarized in table 3.1 with the values that follow from the used specification.

Table 3.1
Effects of patent height with a patent on the basic invention

Patent Height	Nash Equilibrium
$h \in [0, (q - r)/(3q - r))$	$v_s^N = (q - r)/(3q - r)$ $v_L^N = (2q - r)/(3q - r)$
$h \in [(q - r)/(3q - r), 1 - (2\sqrt{(q^3/s)/(3q - r)})]$	$v_1 = v_s^N = (q - r)/(3q - r)$ $v_2 = v_L^N = (2q - r)/(3q - r)$ $v_2 = v_s = h$ $v_1 = v_L = (h + 1)/2$
$h \in (1 - (2\sqrt{(q^3/s)/(3q - r)}), (2q - r)/(3q - r)]$	$v_1 = v_s^N = (q - r)/(3q - r)$ $v_2 = v_L^N = (2q - r)/(3q - r)$
$h \in \langle (2q - r)/(3q - r), 1 \rangle$	$v_1 = v_s = (h(r + q) - r)/2q$ $v_2 = v_L = h$
$h \in [1, \infty)$	$v_1 = 1/2$

9. It is possible that firm 2 does not enter the market for $h < 1$. This is the case from that height on where the isoprofit curve in the new Nash equilibrium also cuts the 45° line with $v_s = v_L$. The profit of the entrant is zero if both product improvements are identical; it is well known that Bertrand competition with homogeneous products leads to zero profits. When lower than 1, this critical patent height is the upper limit.

Perfect Secrecy of the Basic Invention

The scenario described above is an asymmetric variant of the model of Ronnen (1991) and Crampes and Hollander (1991) on minimum quality standards. The secrecy scenario is specific to a context of technical progress and patent protection. If firm 1 can keep its basic invention perfectly secret, it will not patent immediately. By waiting and patenting just before introduction, the patentholder enlarges the protection of his patent, as will be shown below. Price competition remains unchanged compared to section 3.1. The competition in product improvements is now as follows. Firm 1 makes its choice of v_1 in the starting period. Just before the introduction in period $T_1(v_1)$ it takes a *patent on the improvement* v_1 and obtains protection of $[0, v_1 + h]$.¹⁰ The lower limit is 0 because firm 2 will never be allowed to enter with a deterioration. The upper limit is determined by the fact that the patent office requires h to be the minimal inventive step of improvement of the patented improvement v_1 . As soon as firm 1 applies for a patent and enters the market, firm 2 chooses its improvement. Since the information in the patent is public, firm 2 starts from the knowledge level revealed in the patent application. The strategy space of firm 2 is $v_2 \in \langle v_1 + h, 1 \rangle$. A consequence of this structure is that firm 1 will always have the small improvement and firm 2 the large improvement ($v_1 \leq v_2$). A major advantage for firm 1 is that it can make its choice before firm 2. Instead of a Nash equilibrium, the relevant equilibrium concept is therefore a Stackelberg equilibrium, with firm 1 being the leader and firm 2 the follower. The basic Stackelberg equilibrium in improvements is denoted by $\{v_1^S, v_2^S\}$. The three categories of patent heights in the scenario of Stackelberg leadership are summarized in table 3.2. Each category will be explained below.

Table 3.2
Effects of patent height in the case of secrecy of the basic invention

Patent Height	Stackelberg Equilibrium
$h \in [0, (\sqrt{(2q - r)} - \sqrt{(q - r)})/2\sqrt{(2q - r)}]$	$v_1^S = (q - r)/(2q - r)$ $v_2^S = (3q - 2r)/2(2q - r)$
$h \in [(\sqrt{(2q - r)} - \sqrt{(q - r)})/2\sqrt{(2q - r)}, 1/2]$	$v_1^B = 1 - h$
$h \in \langle 1/2, \infty \rangle$	$v_1^M = 1/2$

10. I assume that firm 1 patents the largest improvement available at that time as to rule out the possibility that it takes a patent on the basic invention, while having generated an improvement, in order to be able to have the larger improvement in equilibrium.

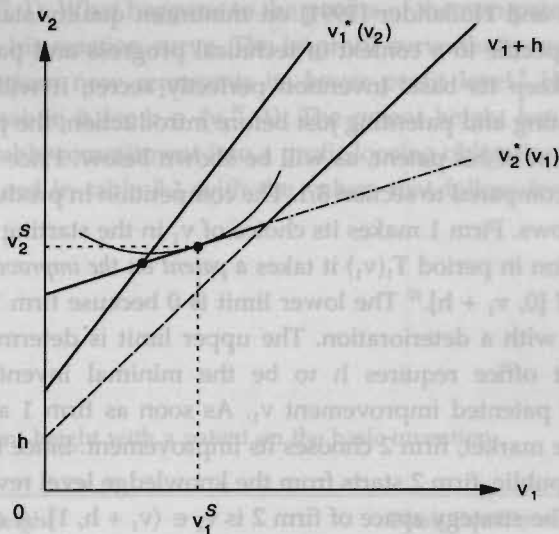


Figure 3.2 Patent on the first improvement

In order to find the basic Stackelberg equilibrium, I first determine the reaction function of firm 2 in this scenario:

$$v_2^*(v_1) = \max((v_1 + 1)/2; v_1 + h) \quad (3.9)$$

The unrestricted part of the reaction function can be used to establish the basic Stackelberg equilibrium as $v_1^S = (q - r)/(2q - r)$ and $v_2^S = (3q - 2r)/(2(2q - r))$. This Stackelberg equilibrium is depicted in figure 3.2 and is only relevant to low patent heights, as will be shown below. The Stackelberg equilibrium for the restricted part, $v_2 = v_1 + h$, turns out to be undetermined because the profits of firm 1 always increase in v_1 , independent of height h . In order to be able to determine which improvement(s) will appear on the market, I define v_1^B as the improvement that blockades the market for firm 2. This means that firm 2's strategy space in improvements is empty (firm 2 will not enter). The blockading improvement here is $v_1^B = 1 - h$. Let the optimal improvement choice of a monopolist be denoted by v_1^M . If the monopoly improvement is larger than the blockading improvement ($v_1^M \geq v_1^B$), firm 1 is better off choosing the monopoly improvement because this also blockades the market and is optimal by definition.

Additionally, if the monopoly profit of firm 1 is larger than its Stackelberg profit, which is the case here, the upper category is established. For $h \in (1/2, \infty)$, firm 1 will be a pure monopolist until doomsday with a product improvement of $v_1^M = 1/2$. For $v_1^M \leq v_1^B$, firm 1 cannot choose the monopoly improvement because firm 2 would then enter. Instead, firm 1 chooses the blockading improvement. Compared to the optimal monopoly profits, the blockading profits decrease if v_1^B is further away from v_1^M , i.e., when patents are lower. The loss in profits due to the strategy of market blockade becomes large to the extent that, from a critical patent height on, firm 1 is better off being a Stackelberg leader in duopoly than a blockading monopolist. Label this height as h'' , defined as $\pi_1(1 - h'') \equiv \pi_1^S(v_1^S, v_2^S)$. The two remaining categories can now be determined. For $h \in [h'', 1/2]$, firm 1 will be a blockading monopolist until doomsday with the blockading improvement $v_1^B = 1 - h$.¹¹ For $h \in [0, h'')$, firm 1 will be Stackelberg leader in duopoly. Proposition 3.2 summarizes the three categories of patent heights.

Proposition 3.2. For $h \in [0, h'')$ the improvements developed in the Stackelberg equilibrium are v_1^S and v_2^S . For $h \in [h'', 1/2]$ the patentholder is a blockading monopolist with an improvement of $v_1^B = 1 - h$; the profits of the patentholder increase in h until monopoly profits are reached for $h = 1/2$. The patentholder continues to be a monopolist for $h \in (1/2, \infty)$ with an improvement of $v_1^M = 1/2$.

3.3 Patent Height and Welfare

The static welfare in the market can be determined in a simple manner because of the linear demand structures resulting from utility function (3.1). The sum of consumer and producer surplus can be taken as an approximation of total welfare. A product improvement generates a level of welfare equal to 3/2 of total profit, both under monopoly and under duopoly. Maximization of welfare is thus equivalent to maximization of joint profits. It can be easily shown that with a patent on the basic invention (scenario [1]), the optimal patent height from a social point of view is given by $h^* = (r + s)/(2s - q + r)$ inside the category $h \in (v_L^N, 1)$ in which the Nash equilibrium with $v_S = v_S^*(h)$ and $v_L = h$ occurs. All heights in the category $[0, h'')$ where a Stackelberg equilibrium with v_1^S and v_2^S occurs are socially optimal in scenario [2].

These welfare conclusions are concerned with the static efficiency once the

11. The critical height h'' in the specification used is not larger than the natural distance between the small and the large improvement in the basic Stackelberg equilibrium ($h'' \leq v_2^S - v_1^S$). The third category, for low patents, then, is $h \in [0, h'')$. Firm 1 will first come up with v_1^S . Afterwards, firm 2 follows with v_2^S .

basic invention is generated. However, a more complete welfare analysis in the context of technical progress would take into account the necessity of an incentive for research. The presence of a basic invention as a starting point is appropriate for competition analysis but not for welfare analysis. How can this research incentive be taken into account? In Klemperer (1990) and Gilbert and Shapiro (1990), it is introduced in the form of a minimal profit V for the firm that undertakes R&D. In the underlying case of patent height, the question arises which firm, 1 or 2, is to be given the incentive V . After all, both firms carry out R&D. In a very specific sense, the profits of the patentholder and the improver stand for special incentives. If the patent office wants to induce *basic research*, it would make the profit of the patentholder equal to V . The patentholder's profits, namely, represent the possible gains from basic research, which is, in a rather limited way, defined here as research that has no direct links with previous research. When patenting the products of basic research, the inventor does not have to take into consideration the minimal steps of improvement. Therefore, if the novelty requirements are more stringent, he only benefits because more protection is provided. He does not face the disadvantage of higher patents, which are harder to overcome. Applied research and development could be defined as research that builds more on previous research. It is in this specific sense that the non-patentholder's profits represent the gains from applied research and development. If the patent office therefore wants to promote *applied research and development*, it should set the height such that the non-patentholder obtains profit V . Depending on the size of V , the optimal height can thus be in any category.

3.4 Conclusions

This chapter has shown that the height of patent protection as determined by the stringency of novelty requirements can affect competition in product improvements. Low patent protection does not affect the natural market equilibrium in product improvements. This result holds regardless of whether the basic invention can be kept secret. In the scenario with perfect secrecy the inventor can already obtain a monopoly position with medium-height patent protection. If secrecy is impossible, the medium category of patent heights is unfavourable to the patentholder since the improver can commit credibly to the more profitable strategy. High patent protection is again favourable to the patentholder. His profits then increase in height, even up to the monopoly level. It should be stressed that these conclusions are only valid under a set of simplifying assumptions which might be restrictive.

Because of the limitations of the model, some possible applications can only

be given with great care. First, it is an empirical fact, although many of their inventions, firms do not patent all of them (see fact a in chapter 1). Because low patent protection does not affect competition, the application costs could be sufficient reason not to patent. Second, in the specific sense explained above, the model might also apply to technology policy. Patent height can be used as an instrument to give relative weights to incentives for basic research vis-à-vis applied research and development.¹² There is a trend in Europe towards uniform patent heights (section 4.3 will outline the details of the European Patent System). Suppose that the uniform patent height of the European Patent System will be the mean of national patent heights. The model predicts that uniform patent height has differential effects on individual countries. Countries with above average high novelty requirements, such as Germany for example, are expected to shift from basic research towards applied research. Countries with low novelty requirements, like for example Portugal, are expected to shift towards basic research.

As explained in the introduction to this chapter, there is a fundamental difference between patent height and patent breadth. While this chapter has purely focused on the height dimension, the interaction with the breadth dimension would also be an interesting object of study. Chapter 5 will examine this interaction; an integrated model, including both breadth and height, will be developed to study optimal strategies for inventing around a patent. Before this integration is established, the next chapter will first study more extensively the dimension of breadth. The relation between patent breadth and the practice of price discrimination will be examined.

12. The emphasis on basic or applied research can be found in national patent policies. Japan, for example, could be considered to have a comparative advantage in applied research and product improvements. Patent protection in Japan is rather low in the sense that each separate claim can be patented (Ordover 1991). It would be premature to conclude that there is a causal relation between low patents and advantage in improvements in Japan, but the co-existence is not illogical. The US and Germany, both countries that can be considered to have a comparative advantage in basic research, have relatively high patents (Ordover 1991).

The tension between patent and competition policy has received much attention in Industrial Organization. It is generally understood that the purpose of patents is to stimulate technical progress. A patent can be necessary to protect an invention because new knowledge that is generated by a firm through research and development (R&D) easily spills over to competitors. A patent provides a temporary monopoly position which enables a firm to extract revenues from the invention and compensate for its R&D expenditures. Although the patent system thus enhances technical progress, static efficiency is disturbed because of the misallocation of resources of the monopolist. This trade-off between static efficiency and technical progress raises tension between patent and competition policy; while patent policy offers monopoly power, competition policy often restricts it. The restrictions on the patentholder's rights, stemming from competition policy, have already been discussed in chapter 2. An important example where restrictions are stringent and the tension is seriously felt is the practice of price discrimination by a patentholder. Schmalensee (1981) and Varian (1985) have examined the general welfare effects of third-degree price discrimination. They have determined rather restrictive conditions for social welfare to increase under price discrimination. Hausman and MacKie-Mason (1988) qualify these results by showing that in scenarios that are typical of new patented products, the tension between patent policy and competition policy is softened. When new markets are served and scale economies are present, social welfare increases under price discrimination.

In this chapter, I will examine whether allowing for price discrimination by a patentholder is a good instrument, from a social welfare point of view, to provide an innovation incentive. An important condition for price discrimination is some market power. The breadth of patent protection, as recently analyzed by Gilbert and Shapiro (1990) and Klemperer (1990), essentially determines the degree of market power of a patentholder. As explained in chapter 2, the breadth

of a patent defines the allowed similarity between imitations and a patented invention. Imitations can be seen as varieties that generate the same gross surplus as the imitated product. Patent breadth can be regarded as an instrument defining the number of protected varieties of a patented product. Like in chapter 3, let us take the example of a new tennis racket (see Klemperer 1990, p.115), and suppose that a new fibre makes it possible to design an oversized tennis racket. The patent protection may then run from 85 to 130 square-inches. Since not all consumers may prefer the same variety, it can be said that patent breadth defines a protected region on the horizontal product spectrum.

More generally, new inventions always contain components that are not new - in some way or another, inventions always build upon other inventions (fact c in chapter 1 provides more backgrounds on this point). A patent office has to decide in two phases of a common patent application procedure (see chapter 2 for an extensive description of a granting procedure) how similar an invention is allowed to be to another invention. First, technical examiners judge whether the claimed invention satisfies the conditions for patentability. One of the conditions is that an invention must be sufficiently novel, which means that it must contain not only a minimum of new components but also a maximum of known components. Second, during the opposition phase, when competitors can make objections against the grant, the patent office must determine when these objections are acknowledged. A patent system that tolerates much imitation during examination and opposition is said to offer narrow protection. If the degree of imitation allowed for is low, patent protection is called broad.

The breadth of patent protection for a new product may differ in two markets, for example because these markets are located in countries applying different patent laws. Consequently, if a firm holds two national patents, its monopoly power differs, *ceteris paribus*, from country to country. The optimal exploitation of both patents then involves price discrimination. This chapter shows that for dynamic efficiency it is better to allow for price discrimination by a patentholder. In addition to the scenario described in Hausman and MacKie-Mason (1988), price discrimination in a dynamic context is thus another exception to the welfare superiority of uniform pricing which is often prevalent.

The chapter is organized as follows: section 4.1 studies the dynamic efficiency under price discrimination and uniform pricing, without reference to patents. In section 4.2 the breadth of patent protection is used to derive the demand functions. Section 4.3 will go into the recent developments in Europe concerning patent protection. Finally, section 4.4 presents the conclusion.

4.1 Price Discrimination and Dynamic Efficiency

Consider a potential inventor who has to be given an incentive to invest in R&D in order to generate a product innovation. The incentive can only be given in the form of a minimum profit level V that can be gained with the new product. Two questions are relevant to the optimal incentive design. First, what is the optimal profit V needed to induce an optimal innovation level? Second, once the optimal V is known, how to provide it? The first question is hard to answer without having some detailed information about the social value of the innovation and its demand and cost characteristics. Fortunately, the second question can be answered separately from the first one. Take the minimum profit level V fixed, as done in Gilbert and Shapiro (1990) and Klemperer (1990). The provision of a positive profit (π) incurs inevitably static welfare losses (WL), i.e., the losses compared to welfare under competitive prices or perfect price discrimination. In order to be able to assess the desirability of price discrimination or uniform pricing, I propose a ratio that indicates how efficiently the profit incentive V can be provided. This ratio is given by the welfare loss per unit of profit, WL/π , and is defined here as the *dynamic efficiency*. If the welfare loss per unit of profit is lower, the innovation incentive can be provided more efficiently. Kaplow (1985) advances a comparable ratio in judging monopolistic practices of patentholders. He proposes to compare the incremental reward to the incremental monopoly loss of a certain practice. If the reward, relative to the loss, of one restrictive practice is higher than that of another, the first practice should be preferred. A similar ratio is also present in Klemperer (1990). The main objective of Klemperer is to find the patent breadth that minimizes the ratio of social cost to profit. Patent lifetime can then be used to bring the profit level of the patentee at the required level V . This chapter will compare two pricing practices of a patentholder. If the profit associated with either practice is insufficient, I assume that a neutral patent policy instrument can bring the profit at the minimum level V .

If the new product of the patentholder is introduced in more than one market, the question arises whether dynamic efficiency improves when the inventor is allowed to apply price discrimination (of the third-degree type)¹, or if he is forced to set uniform prices. I assume that the firm has sufficient market power to be able to practice price discrimination. Obviously, the profits gained from price discrimination are at least as large as the profits gained from uniform pricing, since the option of uniform pricing is also open under price discrimination. With respect to the static welfare losses, one can refer to the well-known

1. See Philips (1989) for an extensive (booklength) treatment of the economics of price discrimination. For a shorter review see Varian (1989).

result in the price discrimination literature that a necessary condition for welfare to increase with price discrimination is that *output* increases, compared to the uniform price setting; a sufficient condition for welfare increase is that the *profitability* of the new output exceeds that of the old output, both valued at the discriminating prices (Varian 1985).

At this point, it can be concluded that dynamic efficiency is enhanced under price discrimination if the static welfare loss with price discrimination is smaller. Important examples are those where a market not served under uniform prices is served under price discrimination. Other examples where static welfare might increase through price discrimination are products whose production technology exhibits declining marginal costs, because of scale or learning economies, but only under the conditions indicated by Hausman and MacKie-Mason (1988).

However, it is not clear whether dynamic efficiency improves or worsens if static welfare does not increase with price discrimination. I will examine this for two markets with linear demands, which are both served under uniform prices. There are at least two reasons why linear demands are examined. First, linearity is often not a bad first approximation of a demand curve. Thus, eventual results are generally applicable. And second, as Schmalensee (1981) and Varian (1985) have shown, in the case of linear demands the output stays the same under price discrimination and static welfare loss is therefore larger under price discrimination. So in the case of linear demands it is indeed not clear what happens with dynamic efficiency.

Consider two independent markets with linear demand curves:

$$x_1(p_1) = a_1 - f_1 p_1; \quad x_2(p_2) = a_2 - f_2 p_2 \quad (4.1)$$

The marginal cost is set equal to zero. Then, it can be easily shown that the optimal prices in market 1 and 2 are $p_1^* = a_1/2f_1$ and $p_2^* = a_2/2f_2$, respectively. The associated profits in each market are $\pi_1(p_1^*) = a_1^2/4f_1$ and $\pi_2(p_2^*) = a_2^2/4f_2$. It can furthermore be shown that the deadweight losses are $WL_1(p_1^*) = a_1^2/8f_1$ in market 1 and $WL_2(p_2^*) = a_2^2/8f_2$ in market 2. The resulting dynamic efficiency is equal to 1/2. Thus, in order to provide one unit of π , welfare losses of $\pi/2$ are born.

Now let us look at the uniform price setting. The relevant part of the joint demand function, where both markets are served, is given by:

$$x_1 + x_2 = a_1 + a_2 - (f_1 + f_2)p \quad (4.2)$$

The optimal uniform price $p_u^* = (a_1 + a_2)/(2(f_1 + f_2))$ yields a total profit for both markets of $\pi_u(p_u^*) = (a_1 + a_2)^2/(4(f_1 + f_2))$. The deadweight loss per market is now

$WL_1(p_u^*) = f_1(a_1 + a_2)^2 / (8(f_1 + f_2)^2)$ and $WL_2(p_u^*) = f_2(a_1 + a_2)^2 / (8(f_1 + f_2)^2)$. The loss in profits (labelled as $\Delta\pi$) due to uniform pricing is indeed non-negative: $\Delta\pi = \pi_1(p_1^*) + \pi_2(p_2^*) - \pi_u(p_u^*) = (a_1f_2 - a_2f_1)^2 / (4f_1f_2(f_1 + f_2))$. Compared to the case of price discrimination, the decrease in welfare loss is $\Delta WL = WL(p_1^*, p_2^*) - WL(p_u^*) = (a_1f_2 - a_2f_1)^2 / (8f_1f_2(f_1 + f_2))$. For $a_1f_2 - a_2f_1 \neq 0$, uniform pricing is indeed welfare-improving ($\Delta WL > 0$). The loss in profits due to uniform pricing is proportional to the lower welfare loss. Dynamic efficiency therefore remains unchanged: $WL(p_u^*) / \pi_u(p_u^*) = WL(p_1^*, p_2^*) / (\pi_1(p_1^*) + \pi_2(p_2^*)) = 1/2$

The absolute welfare loss under linear demands is larger with price discrimination, as was already shown by Schmalensee (1981) and Varian (1985). However, the welfare loss per unit of profit, dynamic efficiency, is equal under price discrimination and uniform prices. In the underlying context where a minimum profit level has to be provided in order to induce R&D, allowing for price discrimination is neutral to requiring uniform prices. This result implies an important reservation toward the prohibition of price discrimination in the competition policy practised in most countries. In the next section, where a 'less partial' study of price discrimination is undertaken, a reserved attitude toward policies forbidding price discrimination is justified even more - it will be shown that dynamic efficiency is enhanced with price discrimination.

4.2 An Address Model of Patent Breadth

In this section, a model is developed which explains the position of the demand curves the patentholder faces, as a function of the protection provided. Suppose that all varieties of the patented product are represented by an address w on the horizontal product spectrum, which extends from 0 to 1. By definition, the address of the patentholder is at 0. Each firm can only have one variety. The ordering in the interval $[0, 1]$ is such that varieties located further away from 0 are less similar imitations²: an imitation at 0 is an exact duplication, and an imitation at 1 is vaguely similar. The breadth b of the patent defines the region protected by the patent to the 'right' of the patentholder: $[0, b)$, with $0 \leq b \leq 1$. Other firms are not allowed in this protected region because they would then infringe the patent. It is assumed that the border of what is judged to be infringe-

2. The patented product can also be assumed to be at the middle of the linear spectrum, or at any point on a circular spectrum, with protection on the 'left' and on the 'right'. Because of the symmetry at both sides, the analysis would not be very different from the one presented here.

ment is precise and known by the patentholder and competitors.³ The claims and description in the patent file and their interpretation by the patent office and the courts define this protected region. It is furthermore assumed that the free range $[b, 1]$ is completely filled up with competitors, for example because imitation is costless. The marginal cost of production of the new product is 0 for both the patentholder and the competitors. The competitors charge a price equal to the marginal cost of 0. The gross surplus provided by these imitations is identical to the original innovation. Each consumer has a most preferred product variety, which is denoted by w^* . I take a continuum of consumers with a uniform distribution of w^* on $[0, 1]$, with density 1. The net surplus U of an individual consumer is given by the following utility function:

$$U = \begin{cases} v - p - td & \text{if the consumer buys} \\ 0 & \text{otherwise} \end{cases} \quad (4.3)$$

where v is the gross surplus, p is the price and d is the Euclidean distance from the consumer w^* to the patentholder at $w = 0$: $d = |w^* - w|$. Since the patentholder is given by $w = 0$ here, the distance d is also uniformly distributed on $[0, 1]$, with density 1. The parameter t is analogous to the transport cost in the Hotelling (1929) model; here, it is a utility penalty associated with consuming a less preferred variety. Consumers located in the free region face no travel costs because there are competitors at each location. The patentholder has to compete with the imitator at the border b for the consumers in the interval $[0, b)$. All consumers are assumed to buy one unit. Let w' represent the consumer who is indifferent to buying the patented product at price p and the product that is supplied at the border b at price $p = 0$. The address of the indifferent consumer is $w' = (tb - p)/2t$. The demand function $x(p; b)$ of the patentholder is:

$$x(p; b) = b/2 - p/2t \quad (4.4)$$

This demand curve is linear, like the ones in the previous section. The linearity follows from the specification of the travel costs, which are linear in distance, and from the uniformity of the distribution of consumers along the line. The breadth b determines for a large part the degree of monopoly power of the patentholder, which is visualized by the position of the demand curve. The demand curve shifts in if the breadth and, consequently, monopoly power decreases. The slope

3. Waterson (1990) examines a patent system where it is not clear beforehand whether the patentholder or the possible infringer wins in court. This uncertainty may affect the patenting decision (see chapter 2 for more details on the Waterson model).

of the demand curve is determined by the unit travel cost. If the unit travel cost is higher, the demand curve is less steep, which means that the demand is less sensitive to price changes.

What are the effects of the institution of patent breadth on static welfare? Maximum static welfare would be obtained with a zero breadth when firms are located all along the line offering the new product at the competitive price $p = 0$. The total welfare would be $W = v$. A positive patent breadth causes two types of welfare losses: (i) the pure *travel costs* (TC) for consumers inside the protected area, who have to travel to the patentholder at the left or to the firm at the border; and (ii) the mark-up-induced-net-travel costs, or loosely defined the *deadweight loss* (DL), due to above-marginal cost pricing by the patentholder.

Consider first the welfare losses due to travelling. In order to focus on the travel costs, I assume for the moment that the patentholder charges the competitive price $p = 0$. The indifferent consumer is $w'' = b/2$. The consumers to the 'left' of w'' buy from the patentholder and those to the 'right' of w'' buy from the competitor at the border. The consumers buying at the border face a welfare loss of $\int_{w''}^b t(q - b)dq = b^2t/8$. Because of symmetry, the same holds for the other side and, thus, total welfare loss due to travelling, or to the absence of products in the protected area, is:

$$TC = b^2t/4 \quad (4.5)$$

The other type of welfare loss DL occurs because of the mark-up of the patentholder on top of the marginal cost. More consumer surplus is lost because, first, consumers that continue to buy from the patentholder have a lower net surplus due to the higher price; and, second, consumers that switch to the competitive substitute at the border have to travel a longer distance. This loss of consumer surplus is only partly made up by the gain in producer surplus. It can be shown that the deadweight loss is given by:

$$DL = p^2/4t \quad (4.6)$$

The profit function $\pi(p)$ of the patentholder can be determined by using demand function (4.4). The optimal price that maximizes this profit function $\pi = x(p)p$ is $p^* = bt/2$. The associated optimal profits are $\pi^* = b^2t/8$. The optimal price of the patentholder p^* makes the deadweight loss in (4.6) equal to $b^2t/8$. The total static welfare loss WL that is due to the institution of patent breadth is illustrated by the shaded area in figure 4.1 and can be easily calculated as $t(w'')^2/2 + t(b - w'')^2/2$, which reduces to:

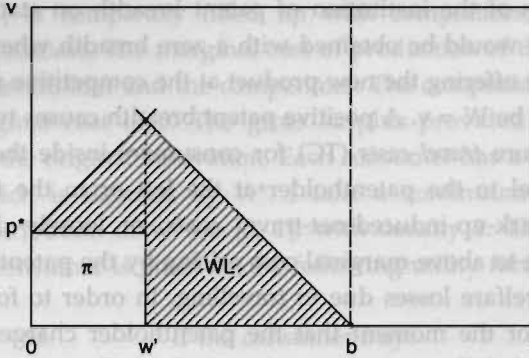


Figure 4.1 Welfare loss due to patent breadth

$$WL = TC + DL = 5b^2t/16 \tag{4.7}$$

Dynamic efficiency in this market is constant and equal to 5/2. It is built up as follows: The travel cost to profit ratio, TC/π^* , is constant and equal to 2. The deadweight loss to profit ratio, DL/π^* , is equal to 1/2 for the optimal price $p^* = bt/2$. Note that without any basics under the demand structure, as in the previous section, only the deadweight loss is caught. The welfare loss which is born because consumers must travel when there are no substitutes, is missing in a 'more partial' analysis. This missed welfare loss turns out to be decisive in the following analysis.

Dynamic Efficiency with Uniform Prices

Consider two markets with different patent breadth, where the protection in market 1 is broader: $b_1 \geq b_2$. The most obvious reason for the difference in breadth is that the markets are situated in countries where there exist different laws concerning patent granting and patent disputes. Chapter 7 explains why two countries which interdependently choose their patent policy, may differ in patent breadth. Another reason might be that the protection provided for the same product within one country differs because of different patent examiners

and judges. The inventor holds a patent in both markets and is located at 0 in each market. Suppose that the patentholder is not allowed to practice price discrimination; he would then have to charge a uniform price p_u in both markets. What is his optimal uniform price? As before, first determine the joint demand function by adding the demand functions. The relevant part of the joint demand function where each demand is positive, is:

$$x(p; b_1, b_2) = (b_1 + b_2)/2 - p/t \quad (4.8)$$

for $0 < p < b_2 + t$. The optimal uniform price can be easily determined as $p_u^* = t(b_1 + b_2)/4$. The corresponding profit is $\pi_u(p_u^*) = t(b_1 + b_2)^2/8$. These price and profits are relevant if both markets are served by the patentholder. However, if there exist substantial differences in breadth in both markets, it may be more profitable for the patentholder to serve only the strong market b_1 . This is because the requirement of uniform pricing has a negative effect on the profit of the patentholder as a result of his deviation from the optimal price p_1^* in each country. This loss in profit can become large to the extent that the optimal profit in the strong market b_1 exceeds the uniform profit in both markets together. The critical value for the narrow breadth is $b_2 = (\sqrt{2} - 1)b_1$. For $b_2 > (\sqrt{2} - 1)b_1$, the uniform profit $\pi_u(p_u^*)$ is larger than the optimal profit in the strong market, $\pi_1(p_1^*)$. The patentholder will only serve the strong market if $b_2 < (\sqrt{2} - 1)b_1$. We have already seen that the dynamic efficiency of providing the R&D incentive in one market is $5/2$. If both markets are served by the patentholder ($b_2 \geq (\sqrt{2} - 1)b_1$), the following holds:

Proposition 4.1. Dynamic efficiency is worse with uniform prices than with price discrimination if the patent breadth in both markets differs. The dynamic efficiency with uniform prices is optimal and equal to the dynamic efficiency with price discrimination if the breadths are identical.

Proof: TC is given for each market by equation (4.5). DL is calculated by substitution of p_u^* in equation (4.6). The dynamic efficiency with uniform prices is then: $WL/\pi_u^* = 1/2 + 4(b_1^2 + b_2^2)/(b_1 + b_2)^2$. Suppose $b_2 = \lambda b_1$, with $\sqrt{2} - 1 < \lambda \leq 1$. We obtain $WL/\pi_u^* = 1/2 + 4(1 + \lambda^2)/(1 + \lambda)^2$. WL/π_u^* decreases in λ and is minimal for $\lambda = 1$. For $\lambda = 1$, $WL/\pi_u^* = 5/2$.

What is the economic intuition behind the proposition? First, note that the dynamic efficiency, as opposed to the static efficiency, is independent of the size of the total patent breadth. It is the *distribution* of the total breadth over the two markets that matters. An indicator for the distribution of the total breadth is λ

(defined as $b_2 = \lambda b_1$, with $\sqrt{2} - 1 < \lambda \leq 1$). The patentholder loses profits if he has to charge a uniform price, because he then deviates from the optimal price in each market. Define the loss of profits as $\Delta\pi \equiv (\pi_1^* + \pi_2^*) - \pi_u^* = \beta(1 - \lambda)^2$, where β is a parameter that takes a positive value. The loss in profits $\Delta\pi$ decreases in λ , that is to say, if the breadth is less unevenly distributed. What is the effect of λ on the two types of welfare losses? The deadweight loss, DL, is smaller under uniform price setting than under, for the patentholder, optimal prices per market. The difference in deadweight loss is $\Delta DL \equiv DL(p_i^*) - DL(p_u^*) = \beta(1 - \lambda)^2/2$. The welfare gain thus increases if the breadths diverge. If λ decreases, the negative effect of loss in profits is compensated at a constant rate by the reduction in welfare loss DL.

The other type of welfare loss that occurs purely because of travelling, TC, is determined for the fixed competitive price of $p = 0$. By definition, it is irrelevant whether the price in a market is the discriminating price p_i^* or the uniform price p_u^* . Hence, the welfare loss TC does *not* compensate for the profit loss. In order to keep TC/π_u^* as small as possible, the profit loss $\Delta\pi = \beta(1 - \lambda)^2$ has to be minimized in λ . This yields $\lambda = 1$. The dynamic efficiency is then $5/2$. Therefore, the reason behind the worse dynamic efficiency of uniform pricing is that the different breadths have full implications for travel cost but not for the patentholder's profits. The patentholder cannot fully exploit the patent breadths, because he is restricted to set a uniform price.

4.3 Economic Implications of Converging Patent Breadth in Europe

Before discussing the issue of patent breadths, I will first give some historical backgrounds of the current patent systems in Europe. Since the end of the nineteenth century, most countries in Europe have a national patent system (Machlup and Penrose 1950). The European patent system is more recent. In the light of the general aim for integration, the European Community has tried to unify national patent systems. The first step towards unification was the Treaty of Strassbourg in 1963. European countries were obliged to conform parts of their national patent laws to statements of the treaty. Although the treaty itself had already been formulated in 1963, it was not until 1980 that it became operative. The second step towards a common EC patent was taken in 1973 with the European Patent Convention (EPC) of Munich. A European patent is centrally granted by the European Patent Office in Munich but regarding the specification it is, to a large extent, subject to national patent laws. A firm chooses the EPC-connected

countries⁴ for which it seeks patent protection. Rather than separate procedures in each chosen country, one procedure is enough for a European patent. The EPC came into operation in most countries in 1980. The logical next step in fulfilling the aim for European unification is the design of a common European patent that applies to all European countries and that is subject to a common European patent law. Such a patent, a Community Patent, was created by the Community Patent Convention of 1975, but is not made operative yet.

Vanhaverbeke and Van Cayseele (1993) show that, by 1990, the EPO has taken over almost the complete 'European patent market' from the national patent systems, the most important reason being the cost of patenting. On average, a European patent is less expensive than three national patents. Therefore, most firms that want protection in three or more European countries apply for a European patent.

It is sometimes argued that a European patent is nothing more than a set of parallel national patents that fall under the national patent laws. This may be true of, for example, the lifetime of the patent, but it is not true in general. The EPC makes exceptions, and certain parts of national patent laws are not valid for European patents. One important exception is the breadth of protection. A European patent has a uniform breadth, equal in all designated countries, while national patents all have different breadths (Pieroen 1988). In some countries, such as Germany and the Netherlands, national patents are broad, while in other countries, such as Great-Britain, national patents are narrow.⁵

Thus, breadth of protection of national patents thus differs. Consequently, if a firm holds various national patents, its market power differs, *ceteris paribus*, from country to country. A conflict then occurs between patent and competition policy. Optimal exploitation of the different degrees of market power involves price discrimination by the patentholder over the different countries. The practice of price discrimination, however, is forbidden by European competition policy in accordance with articles 85-86 of the Treaty of Rome on abuse of market power, and articles 30-36 on the free movement of goods (Goyder 1988). But if forbidding price discrimination, the European community, indirectly, would not respect the rights that countries provide to patentholders. This would be in conflict with article 222 of the Treaty of Rome, where it is stated that "the Treaty shall not prejudice the rules in Member States governing the system of property owner-

4. Fifteen countries have joined the EPC: Sweden, United Kingdom, Germany, Netherlands, Luxembourg, Belgium, France, Spain, Switzerland, Liechtenstein, Austria, Italy, Greece, Denmark and Ireland.

5. The uniform breadth of European patents can be characterized as being intermediate, with a tendency towards narrow interpretation (Pieroen 1988).

ship" (Goyder 1988). Legally, the distinction between the *existence* and the *exercise* of intellectual property rights is used by the Court of Justice of the EC to make judgements about conflicts between patent and competition policy (Friden 1989). The legal concept of existence refers to the conditions for granting patents, such as the novelty requirements. Exercise refers to the effects that granted patents will have, i.e., the rights that the patentholder will be able to exercise, such as price discrimination. The Court of Justice has disclosed, by means of case law, that the existence of national patents is not affected by Treaty rules such as articles 30-36 and 85-86. Countries are thus free to define what conditions they pose on inventions. However, the exercise of the national intellectual property rights is affected by the competition policy of the EC. For this reason, the practice of price discrimination in the case of national patents with different breadths of protection is thus not allowed (according to European competition policy).

The proposition derived in the previous section directly applies to the European context. If a firm holds two national patents with different breadths (for example, a British and a German patent), then, given the fact that European competition policy prohibits price discrimination, the provision of the minimum profit to this firm to make it innovate in the first place is not efficient. The efficiency distortion is less severe if the two breadths are more similar. From a dynamic efficiency point of view, however, it would be best to permit the firm to practice price discrimination over the two countries. The institution of a European patent with uniform breadth, where there is no difference in profits from uniform pricing or price discrimination, is therefore beneficial from the dynamic efficiency viewpoint. These benefits are likely to become effective. As mentioned previously, a firm can choose between a European patent and national patents and due to the lower application cost, a European patent is preferred to three or more national patents.

4.4 Conclusions

A well-known result in the price discrimination literature is that, under linear demand curves, price discrimination worsens static efficiency (Schmalensee 1981, Varian 1985). In this chapter it is shown that, if a firm has to be given a minimum profit level in order to make it innovate, the dynamic efficiency of providing such an incentive is equal under price discrimination and uniform pricing, given linear demand curves. Moreover, when differences in demand functions are caused by patent breadth, the practice of price discrimination of a patentholder is superior to uniform pricing.

Although the simple model presented in this chapter yields powerful results, some limitations of the analysis should be kept in mind. First, only linear de-

mands are examined. Non-linear travel costs and non-uniform distributions of consumers along the product spectrum, leading to non-linear demand curves, should be examined in future research. The same holds for opening possibilities for the patentholder to choose more than one variety within his protected region. Furthermore, the only difference between markets in the above analysis is patent breadth. Of course, there are other factors that cause differences in the demand of the patentholder which may offset the difference in breadth and change the conclusions derived here.

If, however, the focus is purely on breadth, chapter 7 advances a model which explains why national patent breadths may differ. In a two-country model with positive externalities (if one country enlarges its patent breadth, the innovation incentive for the other country's innovators are improved as well), the effects of various asymmetries between both countries are examined. This model can be considered as underlying the assumption of two different national breadths that was made in this chapter.

Before introducing the international setting, the next chapter will develop a model, at industry level, that integrates the dimensions of breadth and height. Various modelling elements from chapter 3 (on height) and chapter 4 (on breadth) are combined into one framework.

Yet the more naive arguments in favour of one or other exclusive right often imply that this alone will be the effect of according the right sought.' (Cornish 1969, p. 18). When we take a closer look at a typical patent procedure (for example, the European procedure discussed in chapter 2), the imperfection of patent protection and the various opportunities for competitors to invent around indeed become clear.

Briefly recapitulating, a patent-granting procedure starts with an application that must contain a specification of the invention. This specification is made up of two parts: a description of the invention, possibly accompanied by drawings, and the claims which indicate the exact protection desired. If an application successfully passes the phases of examination and opposition, a patent is granted, possibly after amendment. The protection provided by the patent is partly determined by the specification and partly by the patent office and the court. The exact protection which was asked for, is written down in the claims. But the court does not have to take these claims literally (in a 'fencepost' system it does); it may interpret the claims in a broader and wider sense (a 'signpost' system). A similar invention that is slightly different from what is written down in the patent specification can also be judged to be covered by the protection. Besides the formulation and interpretation of the claims, the novelty requirements which are used by the patent office in the examination phase, also define the extent of protection of a granted patent. That protection is weak if a current patent can

How can a patentholder exploit his patent if the protection is imperfect and competitors can invent around the patent? As previously exemplified, much more than economists, jurists are aware of the fact that patent protection is imperfect: "To the extent that intellectual property is capable of generating market power, it offers its owner (and his associates) the opportunity to reduce output and raise prices. What it does not bring about is the condition in which the monopolist behaves as though he were the only competitor on the market. Yet the more naive arguments in favour of one or other exclusive right often imply that this alone will be the effect of according the right sought." (Cornish 1989, p. 18). When we take a closer look at a typical patent procedure (for example, the European procedure discussed in chapter 2), the imperfection of patent protection and the various opportunities for competitors to invent around indeed become clear.

Briefly recapitulating, a patent-granting procedure starts with an application that must contain a specification of the invention. This specification is made up of two parts: a description of the invention, possibly accompanied by drawings, and the claims which indicate the exact protection desired. If an application successfully passes the phases of examination and opposition, a patent is granted, possibly after amendment. The protection provided by the patent is partly determined by the specification and partly by the patent office and the court. The exact protection which was asked for, is written down in the claims. But the court does not have to take these claims literally (in a 'fencepost' system it does); it may interpret the claims in a broader and wider sense (a 'signpost' system). A similar invention that is slightly different from what is written down in the patent specification can also be judged to be covered by the protection. Besides the formulation and interpretation of the claims, the novelty requirements which are used by the patent office in the examination phase, also define the extent of protection of a granted patent. The protection is weak if a current patent can

easily be overcome by a small improvement. If the novelty requirements are stronger, patents, once granted, provide more protection.

Thus, there are several opportunities for competitors to circumvent the patent. They can, for example, produce an imitation which does not fall within the interpreted claims, or generate an improvement which fulfils the novelty requirements. Given these opportunities, what is the value of the patent for the patentholder? How much profit can he extract from it? And his competitors? As shown in chapters 3 and 4, the literature on product differentiation, and more specifically the address branch, provides useful tools to examine these questions. Incremental innovations, in the form of imitations or improvements, can be thought of as being differentiations of a certain basic innovation. The competition that occurs between the basic innovator and incremental innovators, or between incremental innovators themselves, can be described by means of address models from the product differentiation literature. This chapter explores the further possibilities of these differentiation models in the context of imperfect patent protection. More precisely, it will work toward a model where patent breadth and height can be studied simultaneously. Given breadth and height of protection, a competitor can determine his optimal strategy of inventing around the patent, which can either be an imitation or an improvement strategy. For eventual policy conclusions (this chapter will only perform positive rather than normative analysis), the effects of the breadth and height standards on the choice of inventing-around strategies may be interesting and will therefore also be examined in this chapter.

The chapter is organized as follows. Section 5.1 briefly defines and formalizes important dimensions and outlines the shape of patent protection. The next sections deal with the question of how the patentholder can exploit this protection. First, section 5.2 studies the profit opportunities for a patentholder facing no competition. Next, scenarios in which the patentholder faces competition from firms that invent around the patent are examined: section 5.3 studies competition from improvement, section 5.4 studies competition from imitation and section 5.5 studies competition from both sources. Finally, section 5.6 discusses the results and provides some concluding remarks.

5.1 The Shape of Protection

As previously shown, the patent dimensions of breadth and height can be formalised by using simple models of product differentiation. Let me briefly repeat the main concepts. First, consider patent breadth. The breadth of a patent defines how similar imitations of a patented invention are allowed to be. Imitations can be seen as varieties of the patented product. Patent breadth can be

viewed as defining the number of varieties of a product innovation that are protected by the patent. Take the example of a new tennis racket (see Klemperer 1990, p. 115 and chapter 3 and 4, where this was also mentioned). Suppose that a new fibre, for example, Du Pont's 'Kevlar' or Akzo's 'Twaron' (see chapter 1), makes it possible to design an oversized tennis racket of, say, 105 square inches. The patent breadth protection on the new racket may then run from 80 to 130 square inches. Since not all consumers may prefer the same variety, it can be said that patent breadth defines a protected region on the horizontal product spectrum. Suppose that all varieties of a patented product can be represented by an address w on the horizontal product spectrum, which extends from 0 to 1. By definition, the address of the patentholder is at 0. The ordering in the interval $[0, 1]$ is such that varieties which are located further away from 0 are less similar imitations. The patent breadth b (≥ 0) protects the range $[0, b]$, where no competitors are allowed. I assume that the border of what is judged to be infringement is precise and known by the patentholder and competitors. The claims and description in the patent file and their interpretation by the patent office and the courts define this protected region.

Next, consider the dimension of height. This dimension indicates how new a new product must be, or to what extent a product must be improved, in order not to infringe a current patent. The stringency of the novelty requirements used by patent examiners mainly determines the height of protection. The dimension shows up most clearly if inventions are related. Similar to chapter 3, I will therefore focus in this chapter on inventions that improve existing products. Height then defines the protection provided by a patent against improvements. Again, take the example of the new tennis racket. Patent height indicates how much a 105 square inch racket must be improved (for example, by using a new, stiffer fibre) in order not to infringe the current racket patent. Since it is reasonable to expect that all consumers prefer an improvement to a product that was improved, patent height can be thought of as defining a protected region on the vertical product spectrum. Improvements of a basic invention can thus be represented by an address v on the vertical product spectrum, which extends from 0 to v^* . The address 0 represents a basic invention and v^* the final improvement possible. Let v be the address, or the improvement level, of the patentholder, which might be the basic invention, given by $v = 0$, or any improvement on $(0, v^*]$. The height h provides protection for innovation v in the interval $[0, v + h]$. Competitors are not allowed here (except, of course, if they have a license permission which is, however, not included here). The lower bound of this protected interval is 0 because an improvement must always be larger than what is publicly known. The patented improvement and the basic invention on which it is based, are the state of the art used by patent examiners as a standard

measure. The upper bound is $v + h$. If the improvement v is patented, the next improvement must be h larger, which is at least $v + h$. In other words, it must fulfil the minimum novelty requirements. If not, the patent on innovation v is infringed.

Finally, there are the dimensions of patent length and width.¹ Patent length needs no further explanation. Although the width dimension is not included in the following analysis, I will give a definition anyway. An invention may contain an idea that can be applied in various products. Sticking to the tennis racket example, the new material the original racket or the improvement was made of could also be used in, for example, fishing rods or squash rackets. The number of applications that are reserved for the patentholder are determined by the patent width. Notice the difference between breadth and width: patent breadth is concerned with the protection on one product spectrum, whereas patent width defines the number of protected product spectra. Because the focus in this chapter is on profits and not on welfare in general, the length of patent protection is less interesting. Calculating the total patentholder's profit is simply a matter of discounting the instantaneous profit, which is determined by breadth and height, over the duration of the patent. Under certain assumptions, the same holds for the width dimension. If the applications are offered on independent markets and the reserved applications are present at the start of the patent without extra cost, it is simply a matter of multiplying the profit per application with the number of reserved applications.²

The total patent protection per reserved application is a function of breadth, height and duration. For one application in one period of the patent life, the protection breadth and height can be represented by a rectangular as in figure 5.1. This protection is valid for a new or improved product with the address 0 on the horizontal product spectrum and the address v on the vertical product spectrum.

1. The survey in chapter 2 distinguished two branches of models on patent breadth. The first branch, where breadth is defined for one market, and the second, which is labelled here as width, for more markets. Patent 'scope' is also used by several authors (Klemperer 1990, Merges and Nelson 1992). I interpret patent scope as a *general* indication of the extent of patent protection. A more precise indication distinguishes the dimensions of breadth, height and width.

2. Relaxing these rather strong assumptions, the non-address branch in product differentiation (for example Dixit and Stiglitz 1977) might be useful to examine patent width.

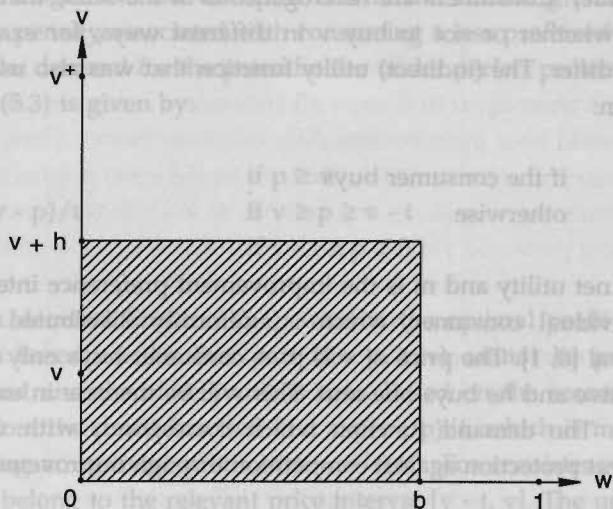


Figure 5.1 The shape of patent protection: height and breadth

5.2 Patent Exploitation by a Pure Monopolist

As mentioned above, this chapter focuses on the breadth and height dimensions of patent protection. The length is assumed to be infinite, although this is a stronger assumption than strictly required (see footnote 12). I will focus on one application, so that the width dimension is also excluded. What are the profit opportunities for a patentholder who enjoys this protection? This section will consider simple benchmark cases where the patentholder does not have to cope with competition: the protection completely covers the interval $[0, 1]$, for $b \geq 1$, and the height interval $[0, v^+]$, for $h \geq v^+ - v$. This will provide some insights into the profit opportunities of the patentholder under perfect protection, as is often assumed, for example in models of patent races. First, I will examine the separate effects of patent height and breadth. After that, the combined effects are studied.

(a) Perfect Height Protection

As explained above, since all consumers prefer a larger improvement to a smaller one, a model of vertical differentiation can be used to examine the height dimension. The innovation of the patentholder has an improvement level v indicating the gross surplus provided by the innovation. This improvement level

v is taken to be exogenous in this section; in the next section v is a deliberate choice of the patentholder. Consumers are heterogeneous in the sense that they evaluate the decision whether or not to buy v in different ways, for example because their incomes differ. The (indirect) utility function that was also used in chapter 3 is taken again:

$$U = \begin{cases} mv - p & \text{if the consumer buys} \\ 0 & \text{otherwise} \end{cases} \quad (5.1)$$

where U is individual net utility and m is the improvement preference intensity parameter of an individual consumer, which is uniformly distributed with density 1 on the interval $[0, 1]$. The price of v is p . A consumer buys only if his net utility is non-negative and he buys one unit. This will be the case in each of the models to follow. The demand function which corresponds with utility function (5.1) and perfect protection against competition through improvement ($h \geq v^* - v$) is:

$$x = 1 - p/v \quad (5.2)$$

For convenience, I set, here and in the rest of the chapter, the marginal production cost equal to 0. The profit function is $\pi = p(1 - p/v)$. The optimal price is $p^* = v/2$, yielding a profit of $\pi^* = v/4$. This is what can be gained with an improvement v , which is perfectly protected against further improvements.

(b) Perfect Breadth Protection

The address w of the patented product on the horizontal product spectrum $[0, 1]$ is 0. The ordering in this interval is such that towards 1, the addresses represent imitations which are less similar to the patented product at 0. At 0 an imitation is thus an exact duplication and at 1 an imitation is vaguely similar. The gross surplus which these imitations provide, however, is identical to the original innovation. As in chapter 4, horizontal differentiation is caught in the following (indirect) utility function:

$$U = \begin{cases} v - p - td & \text{if the consumer buys} \\ 0 & \text{otherwise} \end{cases} \quad (5.3)$$

where d is the Euclidean distance from the consumer w^* to the patentholder at w : $d = |w^* - w|$. Each consumer has a most preferred product variety, which is denoted by w^* . I take a continuum of consumers with a uniform distribution of w^* on $[0, 1]$, with density 1. Since the patentholder is given by $w = 0$ here, the

distance d is also uniformly distributed on $[0, 1]$, with density 1. The parameter t is analogous to the transport cost in the Hotelling (1929) model; here it is a utility penalty associated with consuming a less preferred variety.

The demand for the patentholder under perfect protection ($b \geq 1$) that results from (5.3) is given by:

$$x = \begin{cases} 0 & \text{if } p \geq v \\ (v - p)/t & \text{if } v \geq p \geq v - t \\ 1 & \text{if } v - t \geq p \geq 0 \end{cases} \quad (5.4)$$

Given demand function (5.4), what is the optimal profit for the patentholder? Prices based on the first part (for $p \geq v$) cannot be profit-maximizing since demand and therefore profit is zero. Based on the second part of the demand function, the profit function is $\pi = p(v - p)/t$, which is maximized for $p = v/2$, yielding an optimal profit of $\pi = v^2/(4t)$. For consistency, the optimal price p must belong to the relevant price interval $[v - t, v]$. The upper limit of this price interval, v , is always fulfilled. With respect to the lower limit, p is larger than $v - t$ for $v < 2t$. Demand is non-negative for $v \geq t$. Based on the third part of the demand function, the profit function is simply $\pi = p$; it increases in p in the relevant price range $[0, v - t]$. The optimal price for this range is therefore the highest possible price, $p = v - t$, yielding an optimal profit of $\pi = v - t$. Now compare the optimal profits in both price regimes. For $v < 2t$, the price strategy $p = v/2$ always yields higher profits in the second regime than the alternative price strategy $p = v - t$. In the third regime $p = v - t$ yields more than $p = v/2$.

Summarizing, the optimal price is $p^* = v/2$ for inventions that are small relative to consumer unit travel cost ($t \leq v \leq 2t$). Charging this optimal price for such an innovation yields optimal profit $\pi^* = v^2/4t$. For relatively large inventions ($v \geq 2t$), the optimal price is $p^* = v - t$, yielding an optimal profit of $\pi^* = v - t$. The total market $[0, 1]$ is served then. So, without any competitors in $[0, 1]$, π^* is what can be gained with an invention that is perfectly protected against competition through imitation.

(c) Combining Perfect Height and Breadth Protection

Since an innovation is characterized both by its degree of imitation and its degree of improvement, I need a utility function which includes *both vertical and horizontal differentiation*. The following utility function is a combination of functions (5.1) and (5.3) (see Neven and Thisse 1990 for a similar utility function, and De Palma et al. 1985 and Economides 1986 for less similar functions with two distinguishing characteristics):

$$U = \begin{cases} mv - p - td & \text{if the consumer buys} \\ 0 & \text{otherwise} \end{cases} \quad (5.5)$$

If all consumers had an m equal to 1, (5.5) would be a pure horizontal differentiation model. And if t were equal to 0, or, if all consumers' addresses were at the producer's, (5.5) would be a pure vertical differentiation model. Here, each individual consumer has two characteristics. First, as in (5.1), his improvement preference intensity parameter $m \in [0, 1]$, and second, as in (5.3), the distance $d \in [0, 1]$ between his most preferred variety and the one that is actually offered.

The derivation of the demand function, associated with the patentholder's product at 0 on the horizontal product spectrum and at v on the vertical product spectrum, is less simple now because there are two distributions involved. The distributions of m and d are assumed to be independent; the location of a consumer on $[0, 1]$ does not say anything about the intensity m in which he values improvements. I assume furthermore that the patentholder does not know the combination of characteristics of each consumer.

In order to simplify notations further on, define $\mu \equiv mv$, so that μ is uniformly distributed on $[0, v]$ with density $1/v$; and define $\delta \equiv td$, so that δ is uniformly distributed on $[0, t]$ with density $1/t$. The joint density of μ and δ is $1/(vt)$. The demand for the patentholder is then given by the shaded area in figure 5.2.a and 5.2.b and consists of those consumers with characteristics m and d who have a non-negative net utility: $\mu - \delta \geq p$. For $p \leq v \leq p + t$ (i.e., in the case that even consumers who appreciate improvement most do not buy if their travel costs td are high), the demand is given by:³

$$\int_p^v \int_0^{\mu-p} 1/vt \, d\delta \, d\mu = (v - p)^2 / (2vt)$$

This demand is illustrated in figure 5.2.a by the shaded area in the differentiation box. For $v \geq p + t$, the demand is given by:

$$\int_p^{p+t} \int_0^{\mu-p} 1/vt \, d\delta \, d\mu + t(v - p - t)/(vt) = 1 - p/v - t/2v$$

The shaded area in figure 5.2.b depicts this demand.

3. See Papoulis (1984), chapter 6 for an extensive treatment of joint statistics and functions with two random variables.

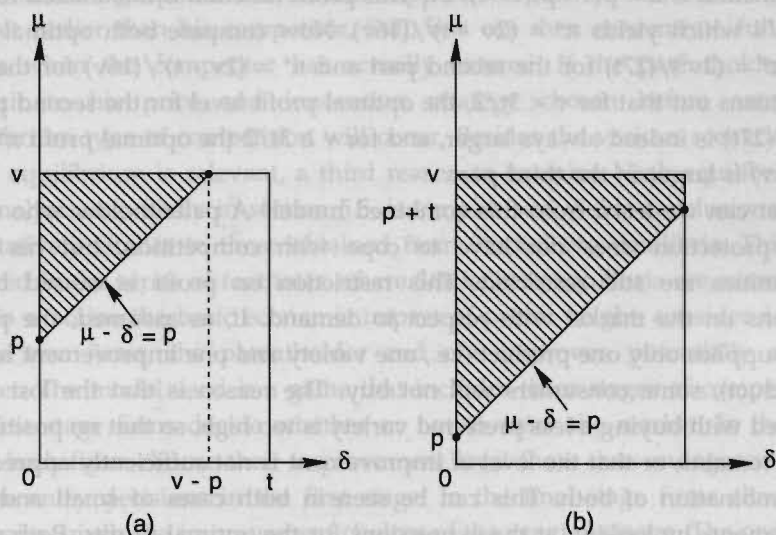


Figure 5.2 Demand for the patentholder as a monopolist

Resuming, the demand function based on the utility function which combines horizontal and vertical differentiation is thus:

$$\begin{aligned}
 x &= 0 & \text{if } p \geq v \\
 (v-p)^2/(2vt) & & \text{if } v \geq p \geq v-t \\
 1-p/v-t/(2v) & & \text{if } v-t \geq p \geq 0
 \end{aligned} \quad (5.6)$$

This demand function is continuous. The second part ($v \geq p \geq v-t$) is strictly convex, the third part is linear ($v-t \geq p \geq 0$) in p . Demand will never be 1 because price cannot be negative; consumers who have a low improvement preference intensity parameter combined with large distance costs do not even buy at zero price. The profit function $\pi(p) = x(p)p$ is also continuous. The optimal price is $p^* = v/3$, for $t \leq v \leq 3t/2$. This part is relevant to inventions which have relatively low improvement levels. The optimal profit is $\pi^* = (2v^2)/(27t)$. For relatively large inventions ($v \geq 3t/2$), the optimal price is $p^* = (2v-t)/4$, yielding a profit of $\pi^* = (2v-t)^2/(16v)$. This can be checked in the following way. Based on the second part of the demand function, the optimal price (for $p \leq v$) is the one which maximizes $\pi = (p(v-p)^2)/(2vt)$, namely $p^* = v/3$. For $v < 3t/2$, p^* belongs to the relevant range $[v-t, v]$. Demand stays non-negative here for $v \geq t$.

The optimal profit with $p^* = v/3$ is $\pi^* = (2v^2)/(27t)$. Based on the third part, the profit function is $\pi = p(1 - p/v - t)/2v$. This profit function is maximized for $p^* = (2v - t)/4$ which yields $\pi^* = (2v - t)^2/(16v)$. Now compare both optimal profit levels, $\pi^* = (2v^2)/(27t)$ for the second part and $\pi^* = (2v - t)^2/(16v)$ for the third part. It turns out that for $v < 3t/2$, the optimal profit level for the second part $\pi^* = (2v^2)/(27t)$ is indeed always larger, and for $v \geq 3t/2$ the optimal profit $\pi^* = (2v - t)^2/(16v)$ is larger in the third part.

What can we learn from this combined model? A patentholder who enjoys perfect protection does not have to cope with competition, but his profit opportunities are still restricted. This restriction on profit is caused by the conditions on the market with respect to demand. If, as assumed, the patentholder supplies only one product (i.e., one variety and one improvement level of the product), some consumers will not buy. The reason is that the lost utility associated with buying a less preferred variety is too high, so that no positive net surplus remains, or that the level of improvement is not sufficiently appreciated, or a combination of both. This can be seen in both cases of small and large improvements by looking at the expressions for the optimal profits. Both profits increase, be it with different speed, if the travel costs of consumers are lower. Abstracting from any research and development cost, the same holds for the improvement level.

5.3 Inventing Around by Competitors: Improvement

Scenarios where the patent does not cover the complete intervals are considered now. I will examine the pure effects of patent height in this section. Competitors can only locate in the unprotected region of the vertical spectrum, $[v + h, v^*]$. In the next sections, the economic effects of breadth and, after that, of breadth and height combined are studied.

But, first, let me briefly comment on an important feature of the analysis in this chapter. Throughout the complete analysis, I will examine games where two firms, the patentholder and a competitor, choose their strategies simultaneously. One can, of course, object that the natural assignment of strategic roles would be for the patentholder to be Stackelberg leader and for the competitor to be the follower. This is because the patentholder can be expected to be first in the new market and therefore able to choose before a potential entrant. Yet the Nash equilibrium analysis (where both firms choose simultaneously) is useful for several reasons. First, if in the pre-patent stage (which is not included here) the eventual patentholder and the eventual competitor were involved in a patent race and this race had a close finish, then it is not likely that the patentholder can make his decisions on entry, price and innovation prior to the other racer. In

such scenarios, which are quite common, Nash equilibrium is more appropriate than Stackelberg equilibrium. Second, suppose the patentholder indeed enters the market earlier than his competitor. Still, one can then concentrate fully on the stages when the competitor has actually entered. If the patentholder cannot commit to his price and innovation strategy chosen before entry, Nash equilibrium type of competition will occur. Besides the various scenarios where Nash equilibrium is relevant, a third reason to examine Nash equilibria is that the analysis, especially in section 5.5, is simpler and yields conclusions that are qualitatively similar to those obtained from Stackelberg equilibria. This can be checked in the series of footnotes where Stackelberg equilibria are examined.

To examine the basic economic impact of patent height, consider a scenario where two firms, the patentholder and one improver, potentially enter the market. The model used is a game that includes three stages of competition. In the first stage firms decide whether or not to enter the market. In the second stage both firms choose their improvement level (now thus endogenous), based on the entry decisions in the first stage. In the final stage firms choose price strategies, given the improvement choices of the previous stage. The order of the stages is determined by the decreasing degree of flexibility of the decisions. Price decisions are more flexible than improvement decisions; improvement decisions, in turn, are more flexible than entry decisions. Only pure strategies will be considered. The solution concept is the subgame perfect equilibrium. As conventional, the game will be solved by backward induction.

Starting with the last stage of price competition, I will first model the demand structure. An individual consumer has the following (indirect) utility function:

$$U = \begin{cases} f + mv_i - p_i & i = 1, 2 \\ 0 \end{cases} \quad (5.7)$$

Firm i has an improvement v_i and charges a price p_i . I take the patentholder as $i = 1$ and the competitor as $i = 2$. Furthermore, I assume an exogenous but natural division of roles: the improvement level of the patentholder is smaller than the improvement level of the competitor ($v_1 < v_2$). Each consumer enjoys an autonomous gross surplus f , which is taken such that the market is completely served in duopoly equilibrium: $f \geq p_1/v_1$. Define m' as the consumer who is indifferent to buying v_1 at p_1 and v_2 at p_2 . This consumer is given by $m' = (p_2 - p_1)/(v_2 - v_1)$. The demand function for the patentholder then is:

$$x_1 = (p_2 - p_1)/(v_2 - v_1) \quad (5.8)$$

The demand function for the improver is the complement: $x_2 = 1 - x_1$. Still assuming zero marginal product cost, the gross profit functions (i.e., without research and development costs) for the patentholder and the improver are:

$$\pi_1(p_1, p_2; v_1, v_2) = p_1(p_2 - p_1)/(v_2 - v_1); \quad (5.9)$$

$$\pi_2(p_2, p_1; v_1, v_2) = p_2(1 - (p_2 - p_1)/(v_2 - v_1))$$

As solution concept for the stage of price choices of 1 and 2, I use the non-cooperative Nash equilibrium. A pair of prices (p_1^*, p_2^*) is an equilibrium if $\pi_1(p_1^*, p_2^*; v_1, v_2) \geq \pi_1(p_1, p_2^*; v_1, v_2)$ and $\pi_2(p_2^*, p_1^*; v_1, v_2) \geq \pi_2(p_2, p_1^*; v_1, v_2)$, for all $p_1, p_2 \geq 0$. Using the gross profit functions in (5.9), the following Nash equilibrium in prices occurs:

$$p_1^*(p_2^*) = (v_2 - v_1)/3; \quad p_2^*(p_1^*) = 2(v_2 - v_1)/3 \quad (5.10)$$

The associated gross profits in equilibrium are:

$$\pi_1(p_1^*, p_2^*) = (v_2 - v_1)/9; \quad \pi_2(p_2^*, p_1^*) = 4(v_2 - v_1)/9 \quad (5.11)$$

The Nash equilibrium always exists.⁴ Firm 2, the improver, is able to undercut the patentholder since, by charging $p_2 < p_1$, he would capture the complete market. But this undercutting price strategy (more precisely given by $p_2 = p_1^* - \epsilon$, where ϵ is small and positive) is always dominated by the price strategy p_2^* in Nash equilibrium (5.10). The gross profits of both the patentholder and the improver, as given by (5.11), increase in distance between improvements in price equilibrium. There is a natural tendency for the improver to improve as much as possible in order to create distance and relax price competition. According to Shaked and Sutton (1982), this is due to the vertical differentiation character of competition in product improvements. Notice that the improver serves the upper segment of the market $[1/3, 1]$, where consumers who appreciate improvement most (large m) are located, while the original innovator serves the lower segment

4. It is assumed here that the patentholder and the improver choose prices simultaneously. A first-mover advantage for the patentholder might be the possibility to choose the price before the improver does. A Stackelberg price equilibrium is then appropriate with the patentholder being the Stackelberg price leader and the improver being the follower. The prices in Stackelberg equilibrium are $p_1^* = (v_2 - v_1)/2$ and $p_2^* = 3(v_2 - v_1)/4$, with corresponding profits of $\pi_1^* = (v_2 - v_1)/8$ and $\pi_2^* = 9(v_2 - v_1)/16$. Notice that these Stackelberg prices and profits of firm 1 and 2 are both higher than the ones in Nash equilibrium.

[0, 1/3]. The improver's larger improvement level enables him to charge a higher price and yet have larger demand. This makes his gross profit (four times) larger than that of the original innovator.

The novelty requirements h that are used by patent examiners determine the minimum distance between both improvement levels: $v_2 - v_1 \geq h$. The novelty requirements can become restrictive for the improver if research and development (R&D) costs of improvements are incorporated. Suppose that increasing R&D leads to a higher gross surplus v being generated by the product innovation, and suppose that this is decreasingly so. A simple innovation cost function which catches the idea of exhausting improvement opportunities is:⁵

$$c(v) = \alpha v^2 \quad \alpha > 0 \quad (5.12)$$

The net profit functions (i.e., gross profit minus R&D costs) then are $\pi_1(v_1, v_2) = (v_2 - v_1)/9 - \alpha v_1^2$ and $\pi_2(v_1, v_2) = (4(v_2 - v_1))/9 - \alpha v_2^2$. A pair of improvements (v_1^*, v_2^*) is an equilibrium if $\pi_1(v_1^*, v_2^*) \geq \pi_1(v_1, v_2^*)$ and $\pi_2(v_1^*, v_2^*) \geq \pi_2(v_1^*, v_2)$, for all $v_1, v_2 \geq 0$. In this case the optimal improvements are independent of each other. The Nash equilibrium in improvements is:

$$v_1^*(v_2^*) = 0; \quad v_2^*(v_1^*) = 2/9\alpha \quad (5.13)$$

The net profit of the improver, $\pi_2^* = 4/(81\alpha)$, is (two times) larger than the net profit of the patentholder, $\pi_1^* = 2/(81\alpha)$.⁶ Because of the exhausting improvement opportunities, the net profit advantage of the improver, relative to the patentholder, is smaller than the advantage in gross profit (which was four times larger, see 5.11).

Three categories of effects of patent height can be distinguished in this simple model. If the patent height is relatively low ($h \leq 2/(9\alpha)$), it does not affect the natural choice of the improver. The protection of the patentholder then does

5. This innovation cost function is in accordance with *Wolff's Law*. "Wolff was a German economist who in 1912 published four 'laws of retardation of progress'. Essentially, he argued that the scope for improvement in any technology is limited, and that the cost of incremental improvement increases as the technology approaches its long-run performance level." (Freeman 1982, p. 216, footnote 2)

6. Based on the Stackelberg *price* equilibrium (see footnote 4), the improvement choices are $v_1^* = 0$ and $v_2^* = 9/(32\alpha)$, yielding profits of $\pi_1^* = 9/(256\alpha)$ and $\pi_2^* = 81/(1024\alpha)$. Since the improvement choices are constant, a Stackelberg equilibrium in *improvements*, based on the Nash *price* equilibrium, does not differ from a Nash equilibrium in improvements.

not enlarge his market power. For intermediate heights ($2/(9\alpha) < h \leq 4/(9\alpha)$), the improver is restricted and the best he can do is to deviate minimally from his optimal improvement and choose $v_2 = h$. Since he must deviate from his optimal choice v_2^* , his profits decrease in h . In this range of height, the profits of the patentholder increase in h because the distance in improvements becomes larger and thus relaxes price competition (compare 5.10 and 5.11). For high patent protection ($h > 4/(9\alpha)$), the improver does not even enter because his profits would then be negative. The patentholder becomes a pure monopolist. These three categories of height effects shift upwards if improvements are less costly to generate (i.e., a smaller α in expression 5.12). Height is then only restrictive for higher values.

The effects of patent height on the competition in product improvements give some indication of how novelty requirements can be used as an instrument of technology policy. If, for example, a government wants to stimulate basic research, it can provide more profit to the patentholder, who generates the product innovation first, by setting stronger novelty requirements. If, on the contrary, it wants to stimulate applied research and development for improvements of products, a patent policy of weak novelty requirements is the more appropriate instrument. Some minimum incentive for basic research always exists because, even if novelty requirements are negligibly weak, the improver freely chooses for some novelty and, consequently, the patentholder has positive profits. It should be stressed here that these conclusions on patent height policy are very premature and based on a simple model. More serious policy implications should be based on richer models.

5.4 Imitation

I will first give some backgrounds of the model of horizontal differentiation which is used here to model breadth.⁷ Hotelling (1929) came to the conclusion that two firms in a linear city will locate both at the centre, in other words, differentiate minimally, if they compete for consumers who have linear travel cost. His conclusion was later corrected by D'Aspremont, Gabszewicz and Thisse (1979), who showed that the price equilibrium at which his statement of minimal differentiation was based, only exists when both firms are located far enough from each other. If located too closely to the other, a firm can gain by undercutting the price and capture the whole hinterland of its opponent. This holds for both firms. As a result, a price equilibrium does not exist when firms are located

7. See Friedman (1983) for a more extensive discussion of horizontal differentiation models and the existence of price equilibrium.

closely to each other. Several solutions have been proposed over time for this problem. I will shortly expose some of them here. Eaton and Lipsey (1975, 1978) use a 'no-mill price-undercutting' assumption as to rule out the inconsistent expectation of the price-undercutting firm. D'Aspremont et al. (1979) specify quadratic transport costs. The price equilibrium then exists but the principle of minimal differentiation turns into one of maximal differentiation. The economic justification of quadratic travel costs, however, is not clear. One could make a counter argument saying that travelling may involve a fixed cost, partially offsetting the marginally increasing variable travel costs, for example because a consumer has to invest a fixed amount in travel equipment, or faces a constant utility penalty if he cannot buy his most preferred variety, independent of distance. Transport costs are then marginally decreasing. Another solution is formulated by Salop (1979). He describes spatial competition on a circle, instead of a line, where a firm has to undercut a competitive price equal to the marginal cost in order to capture the hinterland of a neighbour. Under these conditions the undercutting price strategy is never profitable. Economides (1984) starts from the original Hotelling model and includes a third alternative for consumers, besides the products of both firms. This creates a positive reservation price (consumers have zero reservation prices in the Hotelling model). A firm has to take into account this third alternative when it tries to undercut its opponent. The result is that the range of existence of the price equilibrium widens. Dasgupta and Maskin (1986) study a mixed strategies equilibrium for the Hotelling model rather than the pure strategies equilibria discussed so far. They show that a mixed equilibrium, with intermediate differentiation, does exist for all location choices. Anderson and Neven (1991), finally, examine competition of the Cournot-type instead of the Bertrand-type as is usually done in differentiation models. They show that Cournot competition yields minimal differentiation (spatial agglomeration).

In the following analysis the transport costs are taken linear in distance for at least two reasons. First, as pointed out previously, marginally increasing transport costs are, though technically handy, economically unappealing. Second, the analysis of patent breadth would yield less interesting results if the imitator always want to differentiate maximally, since patent breadth never would then be restrictive. With linear transport cost, there is no tendency towards maximum differentiation. In fact, the imitator will, by assumption, be located at the border of protection at b . In other words, he will always have an imitation that is marginally allowed for by court. If imitation was more costly further away from the patented product, the tendency for minimal differentiation would even be stronger and b would be the free choice of an imitator. In this section on patent breadth and imitation, I will only include the stages of entry and price competition. Location (i.e., imitation) decisions are exogenous.

(a) *One Imitator at b*

I start with the same utility and demand structure as described in section 5.2.b. The patentholder is located at 0, the imitator is located at b . Imitation is assumed to be costless. The market is assumed to be served completely. This is indeed the case if the gross surplus of the product is larger than the highest delivery price of the patentholder or the imitator: $v > \max(p_1 + t, p_2 + bt)$. A consumer buys from firm 1 if $v - p_1 - td_1 \geq v - p_2 - td_2$. The consumer who is indifferent to buying from the patentholder or the imitator is located at address $w = (p_2 - p_1 + tb)/(2t)$. The demand function for the patentholder is:

$$x_1 = \begin{cases} 0 & \text{if } p_1 \geq p_2 + tb \\ (p_2 - p_1 + tb)/(2t) & \text{if } p_2 + tb \geq p_1 \geq p_2 - tb \\ 1 & \text{if } p_1 \leq p_2 - tb \end{cases} \quad (5.14)$$

The demand function for the imitator is $x_2 = 1 - x_1$, by assumption of full market coverage. Firm 1 can undercut firm 2 by charging a price of $p_1 = p_2 - tb - \epsilon$, where, as before, ϵ is small and positive. At this undercutting price, the consumers in the hinterland of firm 2, given by $1 - b$, all buy from the patentholder, making his total demand equal to 1. At a slightly higher price of $p_1 = p_2 - tb$, the total demand of the patentholder is just b , because the consumers in the hinterland $[b, 1]$ do not buy from him but from the imitator. The demand function of the patentholder is thus discontinuous. This makes his profit function also discontinuous. D'Aspremont et al. (1979) have shown that the discontinuity, or, more precisely, the non-quasi-concavity, is a serious problem in the original Hotelling (1929) location model because a Nash equilibrium in prices does not exist when undercutting is profitable. In the underlying model, price undercutting is only possible for one player, the patentholder. As the patentholder is located at the border 0, there is no hinterland to capture for the imitator. The problem of profitable undercutting also arises in this application of the Hotelling model. As will be shown below, a Nash equilibrium in prices only exists if b is sufficiently large.

The first and third part of the demand function (5.14) can be excluded for further analysis. The first part, where demand is equal to 0, can never be optimal for the patentholder; he can always do better by charging a lower price and have some demand and thus profit. The same holds for the imitator regarding part three. Based on the second part of the demand function, the price reaction functions of the patentholder and the imitator are $p_1^*(p_2^*) = (p_2 + bt)/2$ and $p_2^*(p_1) = (p_1 + 2t - bt)/2$. If it exists, the prices in Nash equilibrium are:

$$p_1^*(p_2^*) = t(b + 2)/3; \quad p_2^*(p_1^*) = t(4 - b)/3 \quad (5.15)$$

The associated profits in equilibrium are:

$$\pi_1(p_1^*, p_2^*) = t(b + 2)^2/18; \quad \pi_2(p_2^*, p_1^*) = t(4 - b)^2/18 \quad (5.16)$$

Note that the assumption of the patentholder being located at b is justified; in equilibrium, the imitator, if he were to choose, would want minimal differentiation (note that in 5.16, $d\pi_2/db < 0$, for $b \in [0, 1]$). The location of the imitator, as close as possible to the patentholder and still allowed for by the patent office and by court, is b .

Now I will first determine when these equilibrium prices are in the relevant price interval $p_1^* \in [p_2^* - tb, p_2^* + tb]$ (i.e, the second part of 5.14). Next, I will check when an undercutting price strategy is profitable for the patentholder.

It can easily be shown that the equilibrium prices are in the relevant interval for $b \geq 2/5$. This is the first restriction. Now consider the undercutting price strategy of firm 1. The undercutting price is $p_1 = p_2 - tb - \epsilon$, resulting in a demand of 1 and yielding a profit of $\pi_1 = p_2 - tb - \epsilon$. This undercutting profit must not be profitable: $\pi_1(p_1^*, p_2^*) \geq p_2^* - tb - \epsilon$. This condition can be written as (using 5.16):

$$t(b + 2)^2/18 \geq t(4 - b)/3 - tb - \epsilon \quad (5.17)$$

Condition (5.17) is satisfied for $b \geq 6\sqrt{6} - 14$ (≈ 0.7). So if the patent breadth is smaller than about 0.7, no Nash equilibrium in pure price strategies exists.⁸ Conclusions on patent breadth policy only hold for the range $b \in [0.7, 1]$. The market shares of the patentholder and the imitator are $[0, (b + 2)/6]$ and $[(4 - b)/6, 1]$, respectively. The imitator has the larger profits. Increasing patent breadth levels out market shares, prices and profits. The trade-off in patent breadth policy is between providing sufficient innovation incentive on the one hand, and stimulating, through imitation, both competition, which lowers the deadweight losses of patent monopolies, and diffusion, which enlarges consumer surplus, on the other.⁹

8. For cases where the patentholder can first choose his price, knowing the expected reaction of the imitator beforehand, the Stackelberg equilibrium prices are $p_1^* = t(b + 2)/2$ and $p_2^* = t(6 - b)/4$, yielding profits of $\pi_1^* = t(b + 2)^2/16$ and $\pi_2^* = t(6 - b)^2/32$. The non-undercutting condition is more restrictive in the Stackelberg scenario: only for $b \geq 2\sqrt{41} - 12$ (≈ 0.81) the Stackelberg equilibrium in prices exists.

9. The exact trade-off is relatively easy to determine here. The welfare losses can be split up in two parts. Firstly, there are pure travel costs which occur when both firms charge competitive prices equal to the marginal cost. This welfare loss is $WL_1 = tb^2/4 + t(1 -$

In the remainder of this chapter, three scenarios will be examined where the problem of profitable price undercutting by the patentholder is less restrictive. The first (section 5.4.b) is fragmentation in the free range. As a result of the free competition then emerging in the range $[b, 1]$, prices will be equal to the marginal cost and undercutting this competitive price cannot be profitable. Besides, if the patentholder undercuts the imitator at the border, there will be another imitator waiting next. The second scenario (section 5.5.a) softens the price-undercutting problem by introducing different improvement levels. The patentholder has to overcome the difference in improvements before his price-undercutting strategy becomes effective. If consumers furthermore evaluate improvements differently, as in the third scenario (section 5.5.b), the problem of discontinuous demand and profit functions disappears.

(b) Fragmentation in the Free Range

The interval $[0, b)$ is protected. Suppose that the free range $[b, 1]$ is completely filled up with competitors. These competitors charge the competitive price equal to the marginal cost, which is in this case $p = 0$. Again, I take utility function (5.3). The consumers that are located in the free region face no travel costs because there are competitors at each location. The patentholder has to compete with the imitator at the border b for the consumers in the interval $[0, b)$. I assume that all consumers buy ($v \geq tb$). The demand function for the patentholder is:

$$\begin{aligned} x &= 0 & \text{if } p \geq tb \\ (tb - p)/(2t) & & \text{if } 0 \leq p \leq tb \end{aligned} \quad (5.18)$$

The optimal price, based on the second part of the demand function, is $p^* = tb/2$, which is always in the relevant price interval. The optimal profit is $\pi^* = b^2t/8$.

Some basic policy conclusions can be extracted from this version with free competition. The profit of the patentholder is an indication of the effectiveness of patent policy in providing an innovation incentive. Compared to the presence of only one imitator, patent breadth provides less profit with fragmentation in the

$b^2/2$. The second type of welfare loss occurs because both firms charge above-marginal-cost prices. This 'price-induced-net-travel cost' is $WL_2 = (p_2 - p_1)^2/4t$. After substituting the prices in Nash equilibrium, the total welfare loss is given by: $WL_T = t(31b^2 - 44b + 22)/36$. If a social planner is only concerned with minimizing the total static welfare loss WL_T without providing a minimum profit for a patentholder, it optimally sets a breadth of $b^* = 22/31$. If the social planner aims at providing a minimum profit level for an inventor at minimum static welfare loss, it optimally sets the patent breadth $b^* = 11/14$, which is the one that minimizes the ratio WL_T/π_1^* . For these optimal patent breadths, the Nash price equilibrium exists.

free range. The reason is that the patentholder must compete with an imitator who charges a competitive price equal to the marginal cost. The trade-off for a social planner then goes as follows: discouraging imitations in the free range enlarges the innovation incentive but lowers the consumer surplus in the free range because consumers then have larger travel costs. How exactly this trade-off is best handled is not dealt with here.

5.5 Improvement and Imitation Combined

This section analyses a duopoly where a competitor is in the market with a product that is both an imitation and an improvement of the patentholder's product. As in section 5.3, the stages where firms make entry decisions and choose improvement levels are included and take place before the price competition. Such an analysis serves at least two goals. Thus far, the patent office only requires a minimum level of improvement or an imitation with a minimum distance away from the patent. A product that is both an improvement *and* an imitation provides the opportunity to study the optimal circumventing strategy of a competitor: aside or above the patented product. It offers furthermore the opportunity to investigate the combined effects of the patent policy instruments of breadth and height. First, subsection (a) will examine the case of homogeneous (with respect to the preference intensity parameter m) consumers. After that, subsection (b) will examine the more complex case of heterogeneous consumers.

(a) Homogeneous Consumers

The improvement levels of the patentholder and the imitator differ: $v_1 < v_2$. Consumers are taken to be homogeneous in the sense that they evaluate the difference in improvement levels all in the same way. In the next subsection consumers are taken to be heterogeneous in this sense. Consumers thus only differ in their most preferred varieties on the horizontal spectrum. All consumers are assumed to buy (which is the case for $v_2 > \max(p_2 + tb, p_2 + t(1 - b))$). A consumer buys from firm 1 if $v_1 - p_1 - td_1 \geq v_2 - p_2 - td_2$. The indifferent consumer is located on the horizontal spectrum at $w = (p_2 - p_1 - (v_2 - v_1) + tb)/(2t)$. The demand function for the patentholder is:

$$x_1 = \begin{cases} 0 & \text{if } p_1 \geq p_2 + tb - (v_2 - v_1) \\ (p_2 - p_1 - (v_2 - v_1) + tb)/(2t) & \text{if } p_2 + tb - (v_2 - v_1) \geq p_1 \geq p_2 - tb - (v_2 - v_1) \\ 1 & \text{if } p_2 - tb - (v_2 - v_1) \geq p_1 \geq 0 \end{cases} \quad (5.19)$$

The demand function for the imitator is the complement: $x_2 = 1 - x_1$. The price reaction functions based on the second part of the demand function are $p_1^*(p_2) =$

$(p_2 + bt - (v_2 - v_1))/2$ and $p_2^*(p_1) = (p_1 + 2t - bt + v_2 - v_1)/2$. If it exists, the price Nash equilibrium is:

$$p_1^*(p_2^*) = (t(2 + b) - (v_2 - v_1))/3; \quad (5.20)$$

$$p_2^*(p_1^*) = (t(4 - b) + v_2 - v_1)/3$$

with the associated equilibrium gross profits:

$$\pi_1(p_1^*, p_2^*) = (t(b + 2) - (v_2 - v_1))^2/(18t); \quad (5.21)$$

$$\pi_2(p_2^*, p_1^*) = (t(4 - b) + v_2 - v_1)^2/(18t)$$

From (5.21) we learn that the equilibrium profit of the patentholder increases in patent breadth while the profit of the competitor, who is imitator and improver at the same time, decreases in breadth. This effect of breadth on profits was also detected in section 5.3. As the protected interval on the horizontal spectrum increases, the market power of the patentholder increases with it, at the cost of a loss of market power of the competitor. Another conclusion from (5.21) is that the patentholder's profit decreases in distance between improvements, while the competitor's profit increases. Since consumers are homogeneous in this model, from a vertical differentiation point of view, there is no price-relaxing effect emerging from distance. A larger distance here just means that the competitor has more advantage relative to the patentholder.

What are the conditions under which the Nash price equilibrium as given by (5.20) does exist? Note that the price strategy of undercutting the imitator is not possible if the distance in improvement levels is too large. If $v_2 - v_1 \geq p_2 - tb$, the patentholder can never reach the consumers in $[b, 1]$. The third part of the demand function thus disappears if $p_2 \leq (v_2 - v_1) + tb$. For the equilibrium price p_2^* , this condition becomes: $b \geq 1 - (v_2 - v_1)/(2t)$. For these b 's, undercutting is not possible and the Nash price equilibrium always exists. The condition for non-profitable price undercutting if the third part of the demand function (5.20) exists is more complex. I refer to Van Dijk (1993) for the derivation of this condition. Here, I suppose that the patent breadth (and, thus, the location of the imitator) fulfils the conditions for existence of the Nash price equilibrium. Now look at the stage before the price competition, when both firms choose their improvement levels. The net profit functions contain the gross profit functions (5.21), which are at their equilibrium values after price competition, and the R&D cost function (5.12):

$$\pi_1(v_1, v_2) = (t(2 + b) - (v_2 - v_1))^2 / (18t) - \alpha v_1^2; \quad (5.22)$$

$$\pi_2(v_2, v_1) = (t(4 - b) + v_2 - v_1)^2 / (18t) - \alpha v_2^2$$

This yields the improvement reaction functions $v_1^*(v_2) = (t(b + 2) - v_2) / (18\alpha t - 1)$ and $v_2^*(v_1) = (t(4 - b) - v_1) / (18\alpha t - 1)$. Notice that $dv_1^* / dv_2, dv_2^* / dv_1 < 0$. So if the competitor chooses a larger improvement level, the patentholder can best choose a smaller improvement, and vice versa. We know from (5.21) that the patentholder's gross profits decrease in difference between improvement levels and that those of the competitor increase. The Nash equilibrium¹⁰ in improvement choices is:

$$v_1^* = (3\alpha t(b + 2) - 1) / (6\alpha(9\alpha t - 1)); \quad v_2^* = (3\alpha t(4 - b) - 1) / (6\alpha(9\alpha t - 1)) \quad (5.23)$$

Both improvement levels are assumed to be positive (which is the case for $3\alpha t(b + 2) > 1$). The distance between both improvement levels in Nash equilibrium is:

$$v_2^* - v_1^* = t(1 - b) / (9\alpha t - 1) \quad (5.24)$$

Expression (5.24) has some important implications for patent policy. Broader patent protection makes the 'natural' distance in improvements, $v_2^* - v_1^*$, smaller. If a competitor is forced to choose his imitation further away from the patentholder, he is not prepared to invest as much in R&D for product improvement in order to create vertical distance. As mentioned, the profit of the patentholder directly increases in patent breadth. This direct effect is enforced by an indirect effect of breadth on the improvement choices. The distance between equilibrium improvements, as given by (5.24), decreases in patent breadth b . Because the distance in improvements shortens, broadening protection has this indirect, positive, effect on the profit of the patentholder. The patentholder faces less competition if the other chooses a smaller improvement.

After substitution of the equilibrium improvements (5.23) in the net profit function, it turns out that the equilibrium profit of the competitor is always

10. The patentholder can have a first-mover advantage and choose an improvement level before the competitor. This first-mover advantage does not necessarily show up in the price stage. The Stackelberg equilibrium in improvements, based on the Nash equilibrium prices in (5.20), is given by $v_1^*(v_2^*(v_1)) = 6t(3\alpha t(b + 2) - 1) / z$ and $v_2^*(v_1^*) = (t(18\alpha t(4 - b) + b - 10) / z)$, where $z \equiv 324\alpha^2 t^2 - 54\alpha t + 1$. The distance in improvements is only positive for $b > 4(9\alpha t - 1) / (36\alpha t - 1)$. Along the same lines, the Stackelberg equilibrium in improvements, which is based on a Stackelberg price equilibrium can be determined.

higher than that of the patentholder. More precisely, the difference in equilibrium profits is:

$$\pi_2(v_2^*, v_1^*) - \pi_1(v_1^*, v_2^*) = t(1 - b)(18\alpha t - 1)/(3(9\alpha t - 1)) \quad (5.25)$$

This result, which also occurred in the pure breadth and height scenarios, suggests that R&D competition may better be described by using so called *waiting games* rather than the commonly used *patent races*. In a waiting game for two players, it is better to be second than to be first. Dasgupta (1988) pointed out that the spill-overs from the R&D output of the first firm to the second can be a reason for the profit of the second to be larger. In the model described here, it is even better to be second in the absence of spill-overs in R&D for improvements from the patentholder to the improver. This conclusion must, of course, be qualified in light of the assumptions of the model. It may be weakened, first, by the division of roles imposed in the competition in improvements where the patentholder always has the smaller improvement and, second, by the specification of costly imitation or by the introduction of a time lag for inventing around, during which the patentholder is a monopolist. But, on the other hand, it may be enforced when R&D costs for the original product innovation are incorporated.

(b) Heterogeneous Consumers

Consumers are now also characterized by the parameter m , which is identical for both products and uniformly distributed on $[0, 1]$ with density 1. The indirect utility function (5.5) which combines imitation and improvement is taken again, but now with a constant f which is sufficiently large as to make the product cover the market completely in equilibrium. The demand function of the patentholder is made up of consumers for whom it holds that: $m(v_2 - v_1) - t(d_2 - d_1) \leq p_2 - p_1$. Define $\mu' \equiv m(v_2 - v_1)$ which is distributed uniformly on $[0, (v_2 - v_1)]$ with density $1/(v_2 - v_1)$. The distance from a consumer to the patentholder at 0 is given by d_1 and to the competitor at b by d_2 . What I need now is the distribution of $(d_2 - d_1)$. I first focus on the range $[0, b]$ where $d_1 \in [0, b]$ and $d_2 \in [0, b]$ with density 1. In the range $[0, b]$, it holds for all consumers that $d_2 + d_1 = b$. Therefore, $d_2 - d_1 = 2d_2 - b$, so that $(d_2 - d_1) \in [-b, b]$ with density $1/2$. In order to simplify notations, define $\delta' \equiv t(d_2 - d_1)$ which is distributed on $[-tb, tb]$ with density $1/(2t)$. The joint distribution of μ' and δ' , which are independent, has density $1/(2t(v_2 - v_1))$. The division of the market is given by the line: $\mu' - \delta' = p_2 - p_1$ (see figure 5.3.a and 5.3.b). The consumers who buy from the competitor are located above and to the left of this line. Those who buy from the patentholder are located below and to the right of this line.

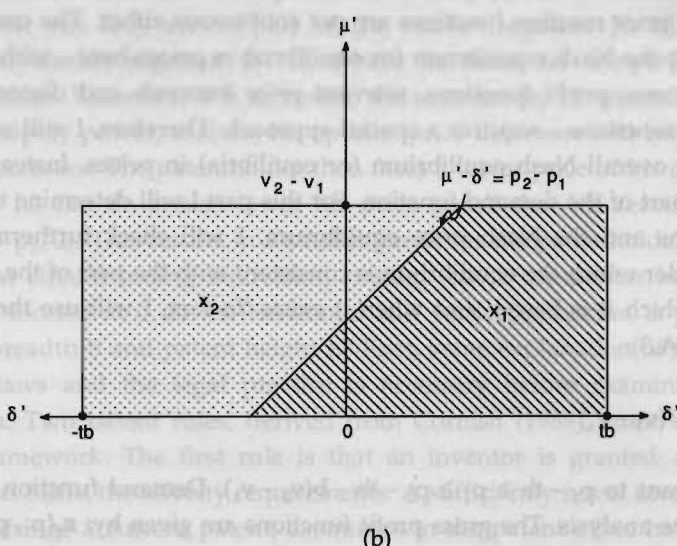
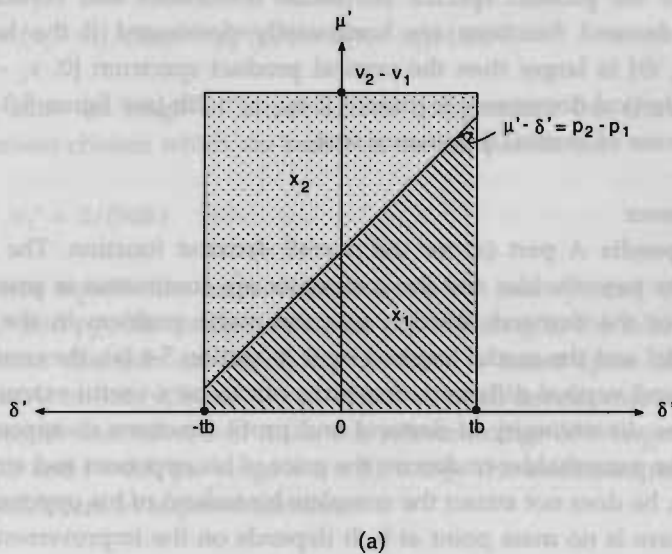


Figure 5.3 Demand in duopoly

Two types of demand functions can be distinguished, depending on the relative size of the product spectra: horizontal dominance and vertical dominance.¹¹ The demand functions are horizontally dominated if the horizontal spectrum $[-tb, tb]$ is larger than the vertical product spectrum $[0, v_2 - v_1]$ (see figure 5.3.b). Vertical dominance is present if $v_2 - v_1 > 2tb$ (see figure 5.3.a). I will work out the case of vertical dominance first.

Vertical Dominance

I refer to Appendix A part (a) for the overall demand function. The demand functions of the patentholder and the competitor are continuous in prices. Since discontinuity of the demand function was the major problem in the original Hotelling model and the model inspired by it in section 5.4 (a), the combination of horizontal and vertical differentiation turns out to be a useful extension. The reason why the discontinuity of demand and profit functions disappears is the following: if the patentholder undercuts the price of his opponent and attracts the consumer at b , he does not attract the complete hinterland of his opponent at the same time. There is no mass point at b . It depends on the improvement preference parameter m of a consumer in $[b, 1]$ whether or not this consumer is attracted.

However, now there is a problem at another level. Although the profit functions are continuous, they are not continuous in first derivatives (in price). Therefore, the price reaction functions are not continuous either. The complexity of determining the Nash equilibrium (or equilibria) in prices here - with six-part demand functions, profit functions, relevant price intervals and discontinuous price reaction functions - requires a special approach. Therefore, I will not try to determine the overall Nash equilibrium (or equilibria) in prices. Instead, I will focus on one part of the demand function. For this part I will determine the gross profit functions and the Nash price equilibrium. I will check furthermore the conditions under which the equilibrium is consistent with the part of the demand function on which it is based, and when it exists. In fact, I will use the middle third part of (A.3):

$$x_1 = b(p_2 - p_1)/(v_2 - v_1) \quad (5.26)$$

which is relevant to $p_2 - tb \geq p_1 \geq p_2 - tb - b(v_2 - v_1)$. Demand function (5.26) is the base for the analysis. The gross profit functions are given by: $\pi_1(p_1, p_2; v_1, v_2) = p_1 b(p_2 - p_1)/(v_2 - v_1)$ and $\pi_2(p_2, p_1; v_2, v_1) = p_2(1 - b(p_2 - p_1))/(v_2 - v_1)$. The corresponding price reaction functions are $p_1^*(p_2) = p_2/2$ and $p_2^*(p_1) = (bp_1 + v_2 -$

11. This typology can be found in Neven and Thisse (1990) and Ireland (1987).

$v_1)/(2b)$, which generate the price Nash equilibrium:

$$p_1^*(p_2^*) = (v_2 - v_1)/(3b); \quad p_2^*(p_1^*) = 2(v_2 - v_1)/(3b) \quad (5.27)$$

Turning into the stage of improvement competition, the (constant) optimal improvement choices which are based on the net profits, are:

$$v_1^* = 0; \quad v_2^* = 2/(9\alpha b) \quad (5.28)$$

Given these equilibrium improvement levels v_1^* and v_2^* , and the equilibrium prices p_1^* and p_2^* , the condition that the prices belong to the relevant interval of the demand function on which the analysis is based (part (iii) of (A.3) in Appendix A) is: $b \leq (\sqrt{36\alpha t + 1} - 1)/(18\alpha t)$. A small development cost parameter α and a small unit travel cost make this relevant range of b larger (for example, for $\alpha = 0.5$ and $t = 0.5$, the range is $b \in [0, 0.48]$). The associated net profits in the Nash equilibrium in improvements are:

$$\pi_1(v_1^*, v_2^*) = 2/(81\alpha b^2); \quad \pi_2(v_2^*, v_1^*) = 4/(81\alpha b^2) \quad (5.29)$$

Note that both profits decrease in patent breadth b . That part of the demand function was chosen where the patentholder only serves consumers in $[0, b]$. The competitor not only serves part of the market segment $[0, b]$, but also the complete market segment $[b, 1]$, where he faces no competition from the patentholder. Therefore, if b increases, the segment $[b, 1]$, where the competitor has monopoly power, shrinks, his optimal price decreases and consequently his profit decreases. The patentholder then faces a lower price of his competitor and the best he can do is also lower his price (recall that his reaction function is $p_1^*(p_2) = p_2/2$). This results in a smaller profit for the patentholder as well.

What circumventing innovation strategy can a competitor best choose if a product located at 0 on the horizontal and vertical spectrum is protected with patent breadth b and patent height h ? First, some explanation of the working of patent laws and the legal practice is necessary before examining the above question. Two patent rules, derived from Cornish (1989), summarize the basic legal framework. The first rule is that an inventor is granted a patent if his invention fulfils the novelty requirements. A sufficiently novel invention can thus never infringe a current patent; examiners in the patent office have checked for this. The rule is independent of the imitation choice of the inventor. The invention is allowed to resemble a current patent to a large extent or not at all, as long as the inventive step relative to the current patent is sufficiently large. The second rule is that a competitor who keeps sufficient distance with his imitation

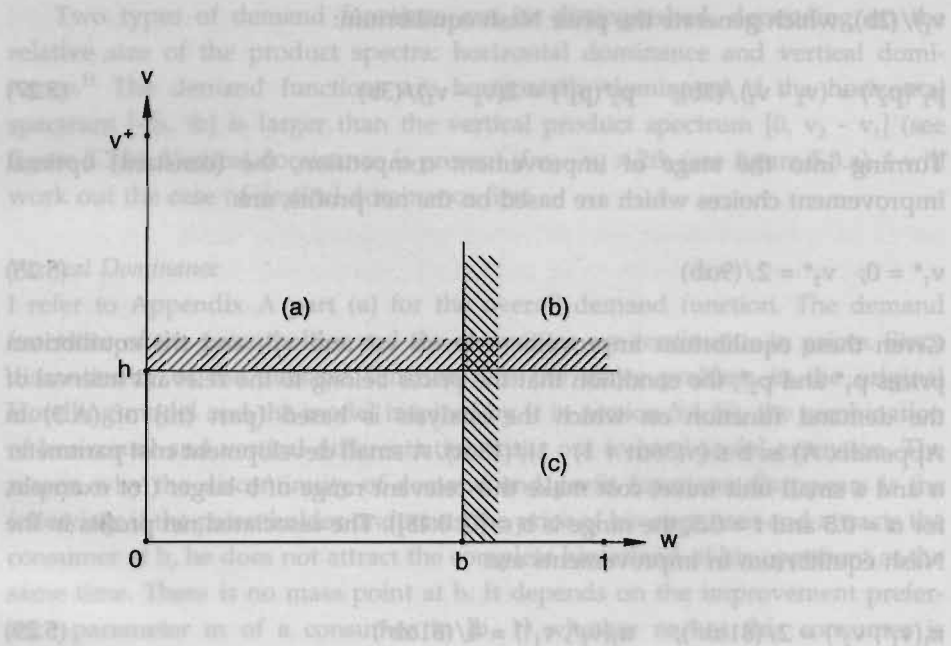


Figure 5.4 Areas of non-infringement

does not have to fulfil the novelty requirements in order not to infringe the current patent. If he does not meet the novelty requirements, he will not be granted a patent himself, but still he will not infringe. These two rules describe three areas in the differentiation box where inventing around is non-infringing. In the first area, (a) in figure 5.4, inventions are sufficient improvements of the basic invention. This feature overrules the fact that they are insufficiently dissimilar imitations. In area (b), inventions are both sufficiently novel and sufficiently different. Inventions in area (a) and (b) are patentable. The inventions in area (c) are not patentable since they are not sufficiently novel. They are, however, tolerated because they are sufficiently different from the basic invention.

Let us now return to the question of optimal strategies for inventing around. A competitor can either choose to focus fully on improvement (in area (a)), or generate a sufficiently distant imitation with some improvement level (in area (b) or (c)).¹² First, consider the *improvement strategy*. If the competitor chooses a

12. Another possible strategy is to wait until the patent has expired and the restrictions imposed by breadth and height withdraw. See Gallini (1992) on this strategy. Here, it becomes clear that the assumption of infinite patents made in the beginning is too

duplication of the patented product at the horizontal spectrum, he has to fulfil the novelty requirements in order not to infringe the patent. Since the patentholder and the improver are then at the same location on the horizontal product spectrum, horizontal distance (difference) plays no role in competition. Recall from section 5.4 that a firm which only differentiates through improvement and not through imitation, has a profit of $\pi_2(v_2^*, v_1^*) = 4/(81\alpha)$ with the improvements $v_2^* = 2/(9\alpha)$ and $v_1^* = 0$ at the market. This improvement choice is allowed for if patent height is low: $h \leq 2/(9\alpha)$. The improver can best choose $v_2 = h$ if the patent height is intermediate ($2/(9\alpha) < h \leq 4/(9\alpha)$). His profit then is $\pi_2(h, 0) = 4h/9 - \alpha h^2$. For high patent protection ($h > 4/(9\alpha)$), however, he cannot make a positive profit with a pure improvement strategy.

The other possible *imitation strategy* of a competitor is to choose an imitation at b which is thus sufficiently different according to patent breadth rules. The imitator's profit then is $\pi_2(v_2^*, v_1^*) = 4/(81\alpha b^2)$ (see 5.29). This net profit decreases in patent breadth. There may be a critical patent breadth which makes this strategy of imitation too costly. It turns out, however, that in the case of vertical dominance, the pure improvement strategy is always dominated, for any combination of breadth and height. To see this take $b = 1$. The imitation strategy then yields $4/(81\alpha)$, which is equal to the unrestricted profit with the pure improvement strategy. For all $b < 1$, the imitation strategy dominates.

Thus, the conclusion is that a competitor who wants to invent around the patent chooses to imitate sufficiently aside the patent and improve somewhat above the patented product. If the improvement level of the circumventer happens to fulfil the novelty requirements, he can obtain a patent himself. Otherwise, he is just tolerated by the patent office and court. At first sight, it is not clear intuitively why the imitation strategy is best in the presence of vertical dominance. One tends to think that inventing around probably occurs where the available space is largest, which in this case is above the patent with an improvement. However, two arguments can be brought forward in favour of the imitation strategy. First, the assumption of costless imitation versus increasingly costly improvement creates a force towards the imitation strategy. Second, as the imitator wants minimal differentiation, the imitation strategy is less attractive if the patent protection is broader. This negative effect of patent breadth cannot become very strong in the case of vertical dominance, where the horizontal spectrum is smaller than the vertical spectrum.

strong. What in fact is assumed is a patent lifetime which is sufficiently long to exclude the 'waiting for the expiration date' strategy.

Horizontal dominance

Next, consider the case of horizontal dominance ($2tb > v_2 - v_1$). The complete demand function of the patentholder (and his competitor) is written out in part (b) of Appendix A. For reasons explained above, I will once again concentrate on one part of the demand function. Part (iii) of A.6 is taken here, since it is in the middle range and linear, which makes it easiest to work with:

$$x_1 = (2bt + 2p_2 - 2p_1 - (v_2 - v_1))/(4t) \quad (5.30)$$

The relevant price interval is $p_1 \in [p_2 - tb, p_2 + tb - (v_2 - v_1)]$. Because the further analysis is technically similar to the above case of vertical dominance, I will not present all calculus details here. The derivation of the price reaction functions, the Nash price equilibrium and the Nash improvement equilibrium can be found in Appendix B.

Only those results which are necessary to determine the competitor's optimal strategy for inventing around are given here. First, the conditions for consistency of the Nash price equilibrium with the part of the demand function used are presented in table B.1 in Appendix B. These intervals form the working space for the further analysis. Second, as previously (check the Nash improvement equilibrium B.7 in Appendix B), patent breadth has a negative effect on the natural distance in improvements. Forcing to keep at greater distance on the horizontal spectrum results in a smaller distance on the vertical product spectrum.

Comparing the strategies for inventing around, an improvement or an imitation strategy, yields richer results than in the case of vertical dominance. In the latter case, the optimal strategy, for any combination of patent breadth and height, was an imitation strategy aside the patent. In the case of horizontal dominance, the optimal strategy depends on both the combination of breadth and height and on the values of the parameters α (the improvement cost) and t (the travel cost). First consider the *imitation strategy*. The net profit level in Nash improvement equilibrium of the competitor who keeps sufficient distance with his imitation and generates some improvement is given by:

$$\pi_2 = (72\alpha t - 1)(12\alpha t(4 - b) - 1)^2 / (144\alpha(36\alpha t - 1)^2) \quad (5.31)$$

The *improvement strategy*, defined as choosing a duplication on the horizontal spectrum and aiming purely at improvement, yields profit of:

$$\begin{array}{ll} \pi_2 = 4/(81\alpha) & \text{if height is not restrictive } (h \leq 2/(9\alpha)) \\ 4h/9 - \alpha h^2 & \text{if height is restrictive } (2/(9\alpha) < h \leq 4/(9\alpha)) \\ 0 & \text{if height is blocking } (h > 4/(9\alpha)) \end{array} \quad (5.32)$$

This profit function is taken from section 5.3.

First, I will compare the imitation profit and the improvement profit if height is not restrictive ($h \leq 2/(9\alpha)$). The optimal strategies for inventing around depend only on patent breadth (and values of α and t). Table 5.1 presents the intervals of patent breadth where the imitation strategy and the improvement strategy are optimal. The lower border is determined by the consistency restrictions derived in Appendix B. From table 5.1. it can be concluded that, like in the previous case of vertical dominance, the imitation strategy dominates the improvement strategy for most patent breadths and parameter values. Only for α 's which are small relative to t (for example, $\alpha = 1/(12t)$ or $\alpha = 1/(10t)$), the improvement strategy can dominate the imitation strategy. Thus, only if the improvement cost is low relative to the consumer travel cost (the travel cost is, in fact, an indication of the imitation cost), inventing above the patent can be more attractive than inventing aside the patent. The improvement strategy actually dominates for large patent breadths.

Table 5.1

Breadth intervals of optimal inventing-around strategies if height is not restrictive

α	Imitation Strategy	Improvement Strategy
$1/(12t)$	[0.500, 0.615]	[0.615, 1]
$1/(10t)$	[0.480, 0.846]	[0.846, 1]
$1/(4t)$	[0.429, 1]	-
$1/(2t)$	[0.414, 1]	-
$1/t$	[0.407, 1]	-
$1/(0.5t)$	[0.403, 1]	-

Next, consider cases where the patent height is restrictive but does not blockade the market ($2/(9\alpha) < h \leq 4/(9\alpha)$). The improvement strategy is even less appealing now. The patent height dictates the improver to deviate from his optimal improvement and choose a larger improvement that does not infringe the patent. Besides breadth, patent height also affects the optimal circumventing strategy now. Figure 5.5 presents the numerical example of $t = 1$ and $\alpha = 1/12$. I chose these values because we know from table 5.1. that for $\alpha = 1/(12t)$, the range where the improvement strategy is optimal is large. Figure 5.5 illustrates the point that for higher patent protection the improvement strategy becomes less

attractive. For unrestrictive height ($h \leq 2.67$) and large breadth ($b \geq 0.62$), the improvement strategy is optimal. For restrictive height, the improvement strategy is only optimal for higher values of patent breadth (for example, for $h = 4.00$ the improvement strategy continues to be optimal for $b \geq 0.93$). From $h = 4.12$ onwards, the imitation strategy is optimal. This illustration is typical of cases where α is small relative to t ; the improvement strategy can be optimal then. For cases where the α 's are larger relative to t , the improvement strategy is never optimal. Finally, if the height is such that it blockades improvements completely ($h \geq 5.33$), an improvement strategy is not even possible. An imitation strategy is always optimal then.

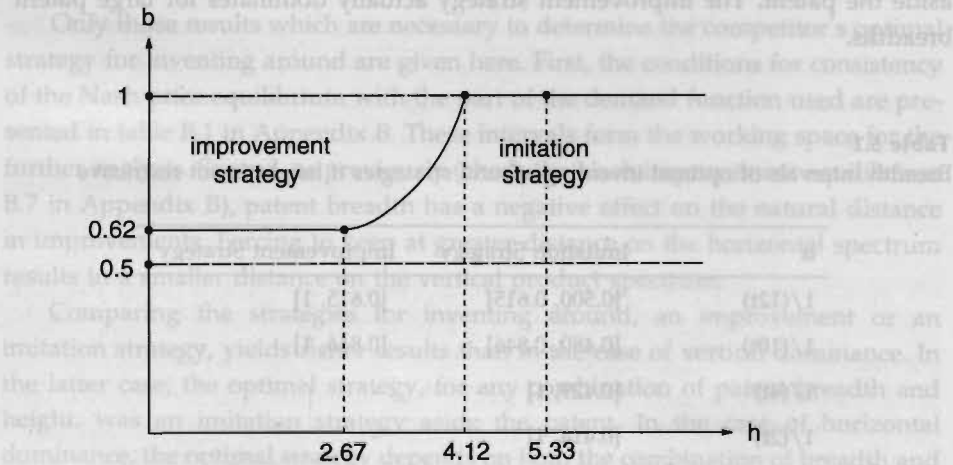


Figure 5.5 Optimal inventing-around strategies (for $t = 1$ and $\alpha = 1/12$)

5.6 Discussion and Conclusion

This chapter has developed a framework, built on address models of product differentiation, in which patent dimensions can be examined. After defining four dimensions of patent protection (length, width, breadth and height), the analysis focuses on two of them: patent breadth and patent height. Sections 5.2 to 5.4 present preliminary models in order to examine the simple bench-mark cases of

monopoly and isolated dimensions. Section 5.5 is the core of this chapter and presents an integrated model of breadth and height. I add a preference parameter to the utility function originally used by Hotelling (1929). The result is a demand structure in which both horizontal and vertical differentiation become visible. So far, the only study which combines these two types of differentiation has been Neven and Thisse (1990). In contrast with the utility function used here, Neven and Thisse specify quadratic transport costs. As d'Aspremont et al. (1979) have shown, quadratic transport costs lead to maximum differentiation in a pure horizontal differentiation model. In order to avoid a tendency for maximum differentiation, which would make the analysis of patent breadth trivial since the competitor would always locate as far away from the patentholder as possible so that breadth would never be restrictive, I have specified linear transport costs, as in the original Hotelling model.

Besides creating the possibility to study patent breadth and height simultaneously, a positive by-product of adding the preference parameter to the Hotelling utility function is that price undercutting strategies, which can prevent the existence of price equilibrium in Hotelling's model, are no longer harmful. The disadvantage, however, is that the analysis of the demand and price reaction functions becomes extremely complex. This has forced me to examine only one part of the demand functions in this chapter. A full analysis incorporating all parts of demand is not included.

In brief, the conclusions are as follows. Patent exploitation by the patentholder is dependent on the standards, breadth and height, set by the patent office and the court. These standards affect the profit opportunities of the patentholder because they determine the strategy space for competitors who surround the patent. Competitors may invent 'above' the patent with an improvement of the patented product, or 'aside' the patent with an imitation. Which circumventing innovation strategy is actually optimal depends on cost and demand parameters as well as on the patent office standards. The exact conditions for choosing an imitation or improvement strategy are not particularly important because they depend on specific modelling assumptions. The result that both innovation strategies can be optimal and are guided by dimensions of the patent system is far more important.

The analysis raises a number of other questions. For example, what are the consequences for the competition in R&D as described in most patent race models? According to Mortensen (1982) there is a socially better allocation of R&D if the winner does not take all but compensates the loser(s). Patent dimensions may be appropriate policy instruments to create such a situation, where the winner of a patent race obtains such a level of protection that the loser

is given enough opportunities to compensate for his R&D expenditures. Furthermore, the purpose of a patent is to provide an incentive for a firm to perform R&D. The incentive is basically given by providing a minimum profit level which can be gained with an innovation. Various mixtures of patent dimensions are possible to provide this minimum profit. The problem then is to find a mixture which causes the smallest static welfare loss. One part of this problem, the effects of patent dimensions on profit, is discussed in this chapter. The other part, the effects on welfare, remain to be studied. Another research question concerns diffusion. True welfare gains occur in the process of diffusion through the economy, with the innovations preferably being supplied competitively. Since patent dimensions also affect the opportunities for competitors, they may have a decisive influence on the speed and amplitude of diffusion. Finally, the patent dimensions discussed may also affect the extent of spill-overs between various innovators. See Steurs (1993) for a model of spill-overs where patent dimensions can play an important role.

Another possible direction based on the analysis in this chapter is empirical research. Some examples of hypotheses that can be formulated on the model explained in this chapter are as follows: first, low and broad protection leads to more improvements, and narrow and high protection to more imitations; and, second, the larger the patent protection provided (higher and broader), the less room there is in the market and, thus, the higher the concentration. However, apart from the problems faced to obtain the appropriate data for testing these hypotheses, the theoretical foundation as advanced in this chapter is rather limited, since only models which focus on one patent and two innovations in a market are analysed, and as such not appropriate for direct empirical testing. A more indirect empirical test, though still based on this chapter, will be carried out in the next chapter, which will study the patenting performance of firms pursuing improvement and imitation strategies.

Appendix A: Derivation of the Demand Curves

(a) Vertical Dominance ($v_2 - v_1 > 2tb$)

The demand for the patentholder (and for the competitor) on the interval $[0, b]$ is made up of different parts:

$$\begin{array}{ll}
 x_1 = 0 & \text{if } p_1 \geq p_2 + tb \quad (\text{A.1}) \\
 (bt - p_1 + p_2)^2 / (4t(v_2 - v_1)) & \text{if } p_2 + tb \geq p_1 \geq p_2 - tb \\
 (b(p_2 - p_1) / (v_2 - v_1)) & \text{if } p_2 - tb \geq p_1 \geq p_2 + tb - (v_2 - v_1) \\
 b - (bt + p_1 - p_2 + v_2 - v_1)^2 / (4t(v_2 - v_1)) & \text{if } p_2 + tb - (v_2 - v_1) \geq p_1 \geq p_2 - tb - (v_2 - v_1) \\
 b & \text{if } p_2 - tb - (v_2 - v_1) \geq p_1 \geq 0
 \end{array}$$

On the interval $[b, 1]$, the difference in travel distance ($d_2 - d_1$) is constant and equal to $-tb$. The market division in the segment $[b, 1]$ is $\mu' = p_2 - p_1 + tb$. In this segment the demand function for the patentholder is:

$$\begin{array}{ll}
 x_1 = 0 & \text{if } p_1 \geq p_2 - tb - b(v_2 - v_1) \quad (\text{A.2}) \\
 (p_2 - p_1 - tb) / (v_2 - v_1) - b & \text{if } p_2 - tb - b(v_2 - v_1) \geq p_1 \geq p_2 - tb - (v_2 - v_1) \\
 1 - b & \text{if } p_2 - tb - (v_2 - v_1) \geq p_1 \geq 0
 \end{array}$$

Depending on the patent breadth, the demand in the segment $[b, 1]$ starts in the third or in the fourth part of the total demand function. Consider the case where it starts in the third part (for $v_2 - v_1 \geq 2tb / (1 - b)$). The total demand over the whole market $[0, 1]$ is:

$$\begin{array}{ll}
 x_1 = \text{(i)} & 0 \quad (\text{A.3}) \\
 \text{(ii)} & (bt - p_1 + p_2)^2 / (4t(v_2 - v_1)) \\
 \text{(iii)} & (b(p_2 - p_1) / (v_2 - v_1)) \\
 \text{(iv)} & (b(p_2 - p_1) / (v_2 - v_1)) + (p_2 - p_1 - tb) / (v_2 - v_1) - b \\
 \text{(v)} & (p_2 - p_1 - tb) / (v_2 - v_1) - (bt + p_1 - p_2 + v_2 - v_1)^2 / (4t(v_2 - v_1)) \\
 \text{(vi)} & 1
 \end{array}$$

Respectively, if $p_1 \geq p_2 + tb$ (i); if $p_2 + tb \geq p_1 \geq p_2 - tb$ (ii); if $p_2 - tb \geq p_1 \geq p_2 - tb - b(v_2 - v_1)$ (iii); if $p_2 - tb - b(v_2 - v_1) \geq p_1 \geq p_2 + tb - (v_2 - v_1)$ (iv); if $p_2 + tb - (v_2 - v_1) \geq p_1 \geq p_2 - tb - (v_2 - v_1)$ (v); and if $p_2 - tb - (v_2 - v_1) \geq p_1 \geq 0$ (vi).

(b) Horizontal Dominance ($v_2 - v_1 < 2tb$)

The division of the market segment $[0, b]$ is:

$$\begin{array}{ll}
 x_1 = 0 & \text{if } p_1 \geq p_2 + tb \quad (\text{A.4}) \\
 (bt - p_1 + p_2)^2 / (4t(v_2 - v_1)) & \text{if } p_2 + tb \geq p_1 \geq p_2 + tb - (v_2 - v_1) \\
 (2bt + 2p_2 - 2p_1 - (v_2 - v_1)) / (4t) & \text{if } p_2 + tb - (v_2 - v_1) \geq p_1 \geq p_2 - tb \\
 b - (bt + p_1 - p_2 + v_2 - v_1)^2 / (4t(v_2 - v_1)) & \text{if } p_2 - tb \geq p_1 \geq p_2 - tb - (v_2 - v_1) \\
 b & \text{if } p_2 - tb - (v_2 - v_1) \geq p_1 \geq 0
 \end{array}$$

On the interval $[b, 1]$, the difference in travel distance ($d_2 - d_1$) is constant and equal to $-tb$. The market division in the segment $[b, 1]$ is $\mu' = p_2 - p_1 + tb$. In this segment the demand function for the patentholder is:

$$\begin{aligned}
 x_1 = 0 & \quad \text{if } p_1 \geq p_2 - tb - b(v_2 - v_1) \\
 (p_2 - p_1 - tb)/(v_2 - v_1) - b & \quad \text{if } p_2 - tb - b(v_2 - v_1) \geq p_1 \geq p_2 - tb - (v_2 - v_1) \\
 1 - b & \quad \text{if } p_2 - tb - (v_2 - v_1) \geq p_1 \geq 0
 \end{aligned} \tag{A.5}$$

The demand on the segment $[b, 1]$ must be added completely to the fourth part of A.4. The total demand on $[0, 1]$ is:

$$\begin{aligned}
 x_1 = & \text{(i)} \quad 0 \\
 & \text{(ii)} \quad (bt - p_1 + p_2)^2 / (4t(v_2 - v_1)) \\
 & \text{(iii)} \quad (2bt + 2p_2 - 2p_1 - (v_2 - v_1)) / (4t) \\
 & \text{(iv)} \quad b - (bt + p_1 - p_2 + v_2 - v_1)^2 / (4t(v_2 - v_1)) \\
 & \text{(v)} \quad (p_2 - p_1 - tb) / (v_2 - v_1) - (bt + p_1 - p_2 + v_2 - v_1)^2 / (4t(v_2 - v_1)) \\
 & \text{(vi)} \quad 1
 \end{aligned} \tag{A.6}$$

Respectively, if $p_1 \geq p_2 + tb$ (i); if $p_2 + tb \geq p_1 \geq p_2 + tb - (v_2 - v_1)$ (ii); if $p_2 + tb - (v_2 - v_1) \geq p_1 \geq p_2 - tb$ (iii); if $p_2 - tb \geq p_1 \geq p_2 - tb - b(v_2 - v_1)$ (iv); if $p_2 - tb - b(v_2 - v_1) \geq p_1 \geq p_2 - tb - (v_2 - v_1)$ (v); and if $p_2 - tb - (v_2 - v_1) \geq p_1 \geq 0$ (vi).

Appendix B: Nash Price and Improvement Equilibrium for Horizontal Dominance

Part (iii) of the demand function A.6 is taken:

$$x_1 = (2bt + 2p_2 - 2p_1 - (v_2 - v_1)) / (4t) \tag{B.1}$$

This part holds for the price range:

$$p_2 + tb - (v_2 - v_1) \geq p_1 \geq p_2 - tb \tag{B.2}$$

The price reaction functions emerging from the gross profit functions $\pi_1 = x_1(p_1)p_1$ and $\pi_2 = x_2(p_2)p_2$ are:

$$p_1^* = (2bt + 2p_2 - (v_2 - v_1)) / 4; \quad p_2^* = (4t - 2bt + 2p_1 + v_2 - v_1) / 4 \tag{B.3}$$

They yield the following Nash equilibrium in prices:

$$p_1^N = (2t(b + 2) - (v_2 - v_1)) / 6; \quad p_2^N = (2t(4 - b) + v_2 - v_1) / 6 \tag{B.4}$$

The associated gross profits are:

$$\pi_1(p_1^N, p_2^N; v_1, v_2) = (2t(b + 2) - (v_2 - v_1))^2 / (72t); \tag{B.5}$$

$$\pi_2(p_1^N, p_2^N; v_1, v_2) = (2t(4 - b) + v_2 - v_1)^2 / (72t)$$

Based on the net profit functions (optimal gross profits minus R&D costs), the impro-

vement reaction functions are:

$$v_1^* = (2t(b + 2) - v_2)/(72\alpha t - 1); \quad v_2^* = (2t(4 - b) - v_1)/(72\alpha t - 1) \tag{B.6}$$

yielding Nash equilibrium improvements:

$$v_1^N = (12\alpha t(b + 2) - 1)/(12\alpha(36\alpha t - 1)); \quad v_2^N = (12\alpha t(4 - b) - 1)/(12\alpha(36\alpha t - 1)) \tag{B.7}$$

The net profit levels in the Nash improvement equilibrium are:

$$\pi_1(p_1^N, p_2^N, v_1^N, v_2^N) = (72\alpha t - 1)(12\alpha t(b + 2) - 1)^2/(144\alpha(36\alpha t - 1)^2)$$

$$\pi_2(p_1^N, p_2^N, v_1^N, v_2^N) = (72\alpha t - 1)(12\alpha t(4 - b) - 1)^2/(144\alpha(36\alpha t - 1)^2) \tag{B.8}$$

The part of the demand function on which the analysis is based is consistent with its relevant price interval as given in (B.2) for:

(a) $p_1^N(v_1^N, v_2^N) \geq p_2^N(v_1^N, v_2^N) - tb;$

(b) $p_1^N(v_1^N, v_2^N) \leq p_2^N(v_1^N, v_2^N) + tb - (v_2^N - v_1^N) \tag{B.9}$

Now, substitute the Nash equilibrium improvements in the Nash equilibrium prices in order to obtain expressions with parameters only. Then substitute these prices and the equilibrium improvements in the conditions (B.9). The lower border (a) reduces to: $b \geq (24\alpha t)/(60\alpha t - 1)$. The upper border (b) reduces to: $b \geq -2(12\alpha t - 1)/(12\alpha t + 1)$. I assume that $12\alpha t - 1 > 0$. Then the upper border (b) is never restrictive for $b \geq 0$. To get an impression of the lower border, table B.1 presents relevant breadth intervals for some combination of α and t .

Table B.1
Consistent breadth intervals

α	Breadth
1/(12t)	[0.500, 1]
1/(10t)	[0.480, 1]
1/(4t)	[0.429, 1]
1/(2t)	[0.414, 1]
1/t	[0.407, 1]
1/(0.5t)	[0.403, 1]
1/(0.1t)	[0.401, 1]
1/(0.01t)	[0.400, 1]

6

Passing the European Patent Office: Evidence from the Data Processing Industry

6.1 Introduction and Theoretical Backgrounds

Not all patents applied for are eventually granted. Applications must pass a granting procedure which examines and selects inventions. As explained in chapter 2, in most systems a patent is only granted for an invention (i) that is new; (ii) involves an inventive step; and (iii) is industrially applicable. This chapter analyzes the success ratio, defined as the proportion of all applications granted a patent, of individual firms in a particular sector: the data-processing industry. What factors determine the success ratio of a firm? As this question is examined for one specific industry, the limitation of analyzing one single industry must be acknowledged. On the other hand, limiting the analysis to a single industry allows one to make some assumptions about relatively 'homogeneous' conditions for all firms. I will not try to answer the question using juridical variables. Thus, interfirm success ratio differences are not, for example, looked for in patent attorneys which may differ in quality. On this point I assume a transparent and competitive market for legal support. The same holds for in-house patent departments of large firms. These departments are assumed to be equally capable.

This chapter examines whether the success ratio can be explained by the type of innovation strategy of the applicant firm. The basic theory underlying this analysis has been described in chapter 5 and will only be roughly sketched here. There are similarities between types of product differentiation and innovation strategies. As Comanor (1967) and De Bondt (1977) have pointed out, an important goal of research and development (R&D) is to achieve product differentiation. A firm which differentiates itself from competitors thereby creates some market power. Although the prospect of market power provides some incentive to perform R&D, the incentive is stronger in the case of patent protection. As explained in chapter 5, a patent defines a protected region around the product, which competitors are not allowed to trespass with their differentiations. This

area of protection might well be larger than the natural distances chosen in free competition. The presence of a patent system therefore provides stronger incentives to invest in product differentiation than in the absence would be the case. The address branch of the product differentiation literature distinguishes two types of differentiation. The first is horizontal differentiation and is relevant if not all consumers agree on the most preferred variety. For example, some consumers prefer a red car while others prefer the same car in white. The second type is vertical product differentiation. The most obvious form of this type is quality. At equal prices, all consumers will prefer a higher to a lower quality.

There is a parallel between these two types of differentiation and two major categories of innovation strategies that can be observed in business practice. The classification here is a simplification of Freeman (1982) to two basic types of strategies. The first is a *pioneering* innovation strategy. Firms that have adopted pioneering innovation strategies allocate more resources to basic research. Their major goal is to generate new or improved products and processes. These innovations are protected primarily by means of patents. Firms pursuing pioneering strategies can rely on patent protection because most innovations they generate are sufficiently novel to pass the patent office. Improvements, but also new products, can be considered as vertical differentiations of existing substitutes. Each consumer will prefer, at equal prices, an improvement to an older version of the product. The other possible strategy is an *imitative* innovation strategy, where more resources are allocated to applied research and development, rather than basic research. The major goal of imitative firms is not to be first with an innovation, but to keep up with the technological leaders in the industry. Their imitations are not basically different but mainly modifications of original innovations. They might be considered horizontal differentiations.

The patenting performance of the two innovation strategies is examined under a single patent system, in this case the European Patent System. Given a consistent patent office policy, pioneering innovation strategies are expected to outperform, with respect to success ratios, imitative strategies. The argument is that a pioneering strategy requires more basic research, which explores novel and unknown technical paths. The novelty requirements of the patent office are therefore met more easily. An imitative strategy aims at modifying and redesigning products with applied research and development. Due to the stronger connection with existing products, the fruits of imitative research pass the patent office less easily.

The underlying study is the first to examine the patent-granting procedure in detail. Griliches (1989) provides some estimates of the waiting times between applications and grants in the US patent office, but he has merely intended to describe the patent procedure. Pakes (1986) and Schankerman and Pakes (1986)

use information from the patent procedure, i.e., the annual renewal fees, in order to estimate the value of patents already granted. They do not, however, examine the possible relation between the value of a patent and the underlying research and development. The study of Mansfield, Schwartz and Wagner (1981) is most related to the present study as it examines innovation and imitation strategies in the presence of patent protection.

The chapter is structured as follows. Section 6.2 describes the two data sets and takes a closer look at the data-processing industry. Section 6.3 is the core of the chapter and presents the estimation results. Section 6.4 examines the success ratios and the waiting times for different countries in the European Patent Office. Finally, section 6.5 offers some concluding remarks.

6.2 Description of the Data

(a) The European Patent Office

As was extensively described in chapter 4, the European Patent Office (EPO), which performs the administrative tasks of the Munich European Patent Convention (EPC), came into force in the second half of 1978. Since then, firms have been able to apply for a European patent. Rather than applying for separate patents in various European countries, a firm may apply for a European patent, which is valid in all EPC-connected countries the firm chooses. Important parts of national patent laws are still valid for the European patent; there is, as yet, no common European patent law. As expected, the EPO has gradually taken over the work of the national patent offices in Europe. The most important reason for this is that a European patent is often less expensive than various national patents. In general, patent attorneys suggest following the European route if firms want protection in three or more European countries, in which case a European application is less expensive (see Vanhaverbeke and Van Cayseele 1993 for more evidence on this issue).

The first data set used here contains information on European patent applications. The set is called 'espace bulletin' and is published by the EPO six times a year. The cd-rom used for this study is the fifth of 1993 and contains all patent applications filed in the EPO from June 1978 to August 1993, plus their procedural status. It is possible, for example, to check whether an application has been withdrawn or granted, which firm is the applicant, from which country the applicant originates and to which technical field the invention belongs, according to the International Patent Classification (IPC). A major advantage compared to other available sets - for example, that of the US patent office - is that it enables one to check whether or not an application is approved, and when.

(b) The Data-Processing Industry

The second data set used here is on the data-processing industry. The data-processing industry is not synonymous with the computer industry. Although the data-processing industry mainly consists of computer firms, part of the telecommunication industry also belong to it. The development of the data-processing industry began around 1951, when the first commercial electronic computer, the UNIVAC I, was installed at the US Bureau of Census. Before, individual inventors in the eighteenth and nineteenth century had made much progress in building calculating machines. But the major impulse towards modern computers was provided by the Second World War, as a result of which electronics and communications techniques had improved considerably. The industry grew rapidly in the 1950s and by 1961 it contained 26 American, 7 British, 3 German, 2 Dutch, 1 Italian and 1 French firms (Williams 1985, Margerison 1978). The rise of the data-processing industry in the 1960s and 1970s is described by Hoffmann (1986). During that period the market was dominated by IBM and the lead position of IBM has continued to the present. IBM still had the first position in the period 1986-1990, which is the period studied in this chapter. However, the dominance of IBM decreased in the 1980s. In 1986 IBM had a market share of 28.3% which decreased to 23.5% in 1990. The four-firm revenue concentration ratio was 41.8% in 1986 and 38.4% in 1990. Thus, of the 4.8% loss in market share of IBM, only 1.4% went to the three firms after IBM. This is largely the result of the entrance of Japanese firms (see Duysters and Hagedoorn 1994 for the internationalisation of the data-processing industry in the 1980s).

The data set used in this study covers the 100 largest firms in the data-processing industry worldwide (by 1990 data-processing revenues) for the period 1986-1990. These data have been collected by a private consulting firm, called the Gartner Group, and are summarized in its report "Yardstick Top 100 Worldwide", September 1991. Table 6.1 exhibits some key figures of the data-processing industry taken from this data set and the EPO. Combining rows 1 and 2 in table 6.1 indicates that the firms active in the market do not tend to concentrate on data-processing: on average, only 31.5% of the corporate revenues originate in the data-processing market. In this data-processing market, the revenues were growing at the high average rate of 17.5% in the years 1986-1988, but slowed to 8.2% for the years 1989-1990. The annual R&D expenditures are 207.7 million dollar on average. The average R&D intensity (R&D expenditures per revenues) is 9.2%.

Table 6.1

Annual key figures of the top 100 data-processing (DP) firms in 1986-1990 (Revenues and expenditures in millions of dollars; employment in # employees)

	Mean	Standard Deviation	Minimum	Maximum
Corporate Revenue	7,171.1	11,998.3	237.8	59,765
DP Revenue	2,259.5	5,853.4	207.8	54,891.2
R&D expenditure DP	207.7	619.7	0	5,715
DP Employment	16,563.7	41,136.8	662	375,587.2
R&D Employment DP	1,853.1	4,181.6	0	35,714.0
# Applications	131.0	327.6	0	1980
# Grants	32.5	91.9	0	619

6.3 Determinants of Success Ratio Differences

Having provided some backgrounds of the data sets and the industry, the analysis can now focus on the specific problem posed in this chapter. First, I will explain the variables that are included in the estimations. The estimation method and the results are presented in the next subsections.

(a) Variables of Specification

The dependent variable is the success ratio of grants to applications for each individual firm in the data-processing industry ('*success*'). The success ratio is calculated for EPO applications in the period 1986-1990. I take all applications of a firm in the IPC classes b41*, g06* and h01* (main and supplementary). These classes fairly well describe the data-processing industry. Next, I look at the number of grants until August 1993 resulting from these applications. An application may have been withdrawn by the applicant, disapproved by EPO examiners, or may still be in process. Most of the 1986-1990 applications are still in the examination process. Thus, the success ratio does not give the absolute proportion of applications being awarded a patent. However, the differences between the provisional success ratios are still expected to provide the necessary information on interfirm differences. A priori, there is no reason to believe that

the waiting times in the EPO are not identical for each firm¹ (section 6.4 will address this issue).

Out of the top 100, 35 firms have not filed an application, so that the success ratio for these firms cannot be calculated. In one series of regressions, I have therefore set the success ratios for these firms equal to zero. The reason for doing so is that the fact itself that a firm does not apply for a patent does provide useful information about its innovation strategy. The absence of applications strongly suggests an imitative strategy. In order to check for the impact of this operation, I have carried out another series of regressions including only firms with 5 or more EPO applications. The sample then reduces to 44 firms. I have taken 5 applications instead of 1 in order to reduce the effect of extreme variables (if a firm applies for only one patent and is granted this patent, the success ratio is 1 and a complete outlier compared to the rest).

The basic theory behind the success ratio was explained in section 6.1. To recapitulate, the hypothesis is that a pioneering strategy, aiming at new or improved products and processes using of basic research, has a higher success ratio than an imitative innovation strategy, which is aimed at imitation and keeping up. The available data set does not contain direct data concerning the distinction between basic and applied research expenditures. Thus, I am looking for indirect variables that indicate the degree of basic research that firms perform, or are expected to perform. Most of the independent variables presented below are inspired by this notion.

The first explaining variable proposed is the ratio of patent applications to R&D expenditures, also known as the propensity to patent (see Scherer 1983). This variable, labelled as '*propens*', is obtained as follows. The total number of applications in the period 1986-1990 is divided by the total R&D expenditures for data processing in that same period. No lag between R&D and patenting is included because there is evidence that most patent applications are filed early in the innovation process (Hall, Griliches and Hausman 1986). I thus obtain the number of applications per million dollar of R&D. It should be stressed that the absolute R&D expenditures will not provide information on the strategies chosen: imitation may cost as much R&D as improvement. The same argument holds for the R&D intensity, defined as the R&D to revenues ratio. The propensity to patent, however, does provide some information on the type of research performed. Following a pioneering strategy, much R&D results in patent applications. Imitative strategies yield less patentable output, and result in a smaller

1. Actually, some country dummies are introduced to deal with possible differences in waiting times.

number of applications per unit of R&D expenditures. A positive sign is therefore expected for 'propens'.

The second and third variables are inspired by the notion that diversification induces more basic research. Nelson (1959) and Arrow (1962) have argued that a more diversified firm can spread risks better and is therefore more inclined to do basic research. Moreover, the output of basic research can be used in more, maybe unexpected, market segments, making the pay-off of basic research higher. As mentioned previously, basic research leads to more successful patenting because the novelty requirements are met more easily. In this study, I distinguish internal from external diversification. By *internal*, I mean the diversification within the data-processing market. A measure for the internal diversification is the number of DP market segments in which the firm has been active in the period 1986-1990, divided by the total number of segments during that period. Label this variable as '*divers*'. I expect a positive sign for '*divers*'. The segments within the data-processing market are mainframes, superminis, mini-computers, microcomputers, cad/cam/cae, peripherals, data communication hardware, software, maintenance, services and other. Over the period (5 years) concerned, the total number of segments was 55. The variable '*divers*' is thus continuous by approximation. The other variable is associated with the *external* diversification of a firm. This variable is defined as the total revenues from the data-processing market divided by the total corporate revenues, for the period 1986 to 1990. Label this variable as '*specialis*'. The same arguments as in the case of internal diversification can be used to justify the inclusion of '*specialis*'. A negative sign is therefore expected. However, it is felt that, since the spill-overs within the market are probably stronger than those coming from outside the market, '*specialis*' will play a less important role than '*divers*'.

The fourth explaining variable is proposed to describe the main focus of a firm, which can either be on innovation or on marketing. Define the variable '*market*' as the ratio of Sales and Marketing employees to R&D employees in data processing. A firm that stresses R&D and wants to compete primarily through innovation is expected to know the technical fields and current patents better and thus to have a higher success ratio. In contrast, a firm that focuses on sales and marketing as a means to compete, knows less of the existing technologies in the market, has less patenting experience, and is thus expected to achieve a lower success ratio. A negative sign is expected for '*market*'. The next variable is concerned with the personnel in the research department. Let '*rdequip*' be defined as the R&D expenditures per R&D employee. The idea behind this variable is that researchers involved in basic research have more equipment at their disposal. So a positive sign is expected for this variable.

The final two variables are suggested to catch some fixed effects. The first is a country dummy. Each country has a different national patent system, which may affect the strategies of firms as well as the experience with patent offices in general. Furthermore, the waiting times of processing in the EPO may differ between the countries of origin of the applicants. Some countries may have to wait longer than other countries. The data-processing firms are divided into three groups by region of origin of their headquarters: United States ('US'), Japan ('JP') and Europe ('EC').² The regressions will include two of the three dummies ('US' and 'JP').

The second dummy variable included indicates the primary market segment (by revenue) in which the firm is active. The 11 segments are regrouped into 3. The first ('computer') contains the core of the data-processing industry with the mainframe, supermini, minicomputer and microcomputer segments. The second group ('noncore') contains the cad/cam/cae, peripheral and data communication segments. Finally, the third group contains the software, services and maintenance segments ('support'). I expect that 'computer' and 'noncore' play a larger role than 'support', because this last group contains such segments as service and software, which can rarely be protected by patents.

(b) Estimation Method

Ordinary Least Squares (OLS) estimations can cause two problems in the context studied here. First, the dependent variable 'success' cannot be smaller than 0 or greater than 1 and is thus doubly truncated: values smaller than 0 must be set equal to 0 and values greater than 1 must be set equal to 1 (see Maddala 1983 for more details on truncated and censored variables). The second problem is caused by the many zero observations for the variable 'success', especially in the large sample series (A). These zero observations can 'force' the constant of the estimated OLS line too much toward zero. Instead of OLS, a two-limit (also known as 'doubly truncated') Tobit estimation model is more appropriate for two reasons. First, the Tobit model deals properly with the lower limit of 0 and the upper limit of 1 of the dependent variable 'success'. Second, the Tobit model also takes into account the many zero observations of the 'success' variable (which is particularly important in the (A) series). In fact, the two-limit Tobit model provides three estimates: (i) of the chance that the success ratio is zero, i.e., whether or not a firm is granted at least one patent (this is a probit); (ii) of the chance that all applications are granted a patent (another probit); and (iii) of the

2. The US group contains 62 firms (including one Canadian firm), the Japanese group 19 firms (including three Taiwanese firms) and the European group 19 firms.

success ratio, given that it is not equal to 0 or 1 (regression part).

(c) Results

Two series of Tobit estimations are carried out. Table 6.2 summarizes the results: in one series (A in table 6.2) the success ratio of firms that have not filed any applications in the EPO in the period 1986-1990, is set equal to 0. In the other (B), firms with fewer than 5 applications are expelled. The major difference between the A-series and the B-series of estimations is the constant term. In the A-series, the constant has a negative sign (all significant at a 0.5% level) indicating the presence of a threshold. Since in the B-series the constant has a positive sign, the firms with fewer than 5 applications face the threshold and cause the negative sign of the constant in the A-series. According to estimation A (iv), for example, where 'divers' is the only explaining variable, a minimum level of internal diversification is required (in this case $0.3309/0.7018 \approx 0.47$, i.e., on average a firm has to be active in about 5 of the total 11 market segments) in order to be granted a patent.

I will discuss the estimation results per variable, starting with those that perform relatively badly. The first is 'rdequip'. The equipment per R&D employee does not have the expected effect on the success ratio. Both in A (i) and in B (i), the variable 'rdequip' does not have a positive sign. Since both coefficients are statistically not significant at the 10% level, no conclusions can be drawn about the effects of the R&D expenditures per R&D employee on the success of passing the EPO. The second variable is 'market'. This variable, which was constructed to catch the main focus of a firm, on marketing or on innovation, does not perform well. 'Market' does not have the expected sign in estimation A(i), while it does have it in B (i). In both cases, however, the coefficient is not significant at the 10% level. Firms pursuing pioneering strategies apparently take into account the marketing of their products as much as firms pursuing imitative strategies. Thirdly, the variable 'specialis', included as an indication of the external diversification of a firm, has the expected negative sign in both A (i) and B (i). But this variable is only significant at the 10% level in A (i) and not in B (i). Thus, external diversification, for spreading risks and absorbing spill-overs from outside the data-processing market, does not seem to be very important in explaining the success ratios.

Table 6.2
Results of the two-limit Tobit estimations

Series	A				B		
	(i)	(ii)	(iii)	(iv)	(i)	(ii)	(iii)
Constant	-0.4039 (-3.007)	-0.4751 (-4.546)	-0.3403 (-4.151)	-0.3309 (-5.139)	0.2068 (2.167)	0.1798 (2.070)	-0.0218 (-0.360)
Divers	0.6458 (4.944)	0.6818 (5.321)	0.6692 (7.629)	0.7018 (7.521)	0.2922 (2.858)	0.2779 (3.299)	0.3574 (3.394)
Propens	0.1635 (3.314)	0.1992 (4.029)	0.2165 (3.005)	-	0.0306 (0.867)	0.0616 (2.819)	-
US	0.0450 (0.562)	0.0216 (0.288)	0.0066 (0.089)	-	-0.1728 (-2.554)	-0.2083 (-3.586)	-
JP	-0.0932 (-1.155)	-0.0952 (-1.168)	-0.0664 (-0.759)	-	-0.1703 (-2.865)	-0.1794 (-3.015)	-
Computer	0.1836 (1.569)	0.1056 (0.907)	-	-	0.0273 (0.307)	-	-
Noncore	0.2558 (2.691)	0.2313 (2.472)	-	-	0.0275 (0.412)	-	-
Specialis	-0.1416 (-1.525)	-	-	-	-0.0990 (-1.235)	-	-
Market	-0.0183 (-0.586)	-	-	-	-0.0003 (-0.024)	-	-
Rdequip	-0.2581 (-0.871)	-	-	-	-0.1497 (-0.718)	-	-
Mean log-lik.	-19.055	-20.287	-24.371	-29.091	21.011	19.540	11.603
Std. error	0.2248	0.2282	0.2408	0.2542	0.1113	0.1147	0.1421
# Cases	100	100	100	100	44	44	44

Note: The dependent variable is 'success'. Numbers in parentheses are t-statistics. The standard errors used to calculate these t-statistics are heteroscedastic-consistent estimates.

2. The US group contains 52 firms (including one Canadian firm), the Japanese group 19 firms (including three Taiwanese firms) and the European group 19 firms.

The results concerning the dummy variables for market segments and countries are mixed. The segment dummies 'computer' and 'noncore' in equations A (i) and (ii) have the expected positive sign. But only 'noncore' firms, and not 'computer' firms, have significantly higher success ratios than 'support' firms (at the 5% level). However, this conclusion does not hold for the B (i) estimation: both segment dummies are not significant at the 10% level. The country dummy variables 'US' and 'JP' perform very well in the B (i) and B (ii) estimations. They both have a negative sign, statistically significant at the 1% level. So US and Japanese firms have lower success ratios than European firms. Section 6.4 studies these country differences in more detail. The country dummies do not perform well in the A-series; in all of the estimations both dummies are not significant at the 10% level.

The next variable, 'propens', performs well in general; it has the expected sign and is statistically significant at the 0.5% level in all regressions but B(i). Therefore, it can be concluded that a high propensity to patent (applications to R&D expenditures ratio) of a firm in the data-processing industry is indeed an indication of the degree of fundamental research and as such associated with good patenting performance.

Finally, the variable for internal diversification, 'divers', performs very well. In each of the seven estimations (A-series and B-series), this variable has the expected sign and is statistically highly significant. Notice that in the estimations with 'divers' as the only explaining variable, i.e., A (iv) and B (iii), the standard error of the estimation is only a small fraction larger than in the estimations including all explaining variables (A (i) and B (ii)). So 'divers' is the most important variable in the estimations.

Since larger firms are more diversified in general, one might object that *firm size* does the job rather than the degree of diversification. In order to check for that, I have defined 'dprev', the revenues in the period 1986-1990, as an indication of size. The variables 'divers' and 'dprev' indeed correlate significantly. The two-limit Tobit estimation (necessary because 'divers' also lies between 0 and 1, inclusive) for the large sample (A-series, 100 cases) is given by: $\text{divers} = 0.3631 (14.840) + 1.5635\text{E-}6 (3.918) \text{dprev}$ (t-values in parentheses). The relation is less strong for the small sample (B-series, 44 cases): $\text{divers} = 0.5052 (13.566) + 2.4178\text{E-}6 (1.849) \text{dprev}$. However, 'dprev' performs worse than 'divers' in explaining the success ratios. For the large sample, the Tobit method results in the following estimation: $\text{success} = -0.1607 (-2.353) + 1.7039\text{E-}6 (5.049) \text{dprev}$. The sign of 'dprev' is the expected one and significant at the 0.1% level. However, compared to estimation A (iv), with 'divers' being the only explaining variable, the t-statistic of 'dprev' in the above estimation is smaller (5.049 compared to 7.521 in A (iv)), and the standard error of the estimation is larger (0.3842

compared to 0.2542 in A (iv)). Therefore, for the large sample it can be concluded that internal diversification has more influence than size in explaining the success ratios. For the small sample this conclusion holds even stronger. The Tobit model yields the following estimation: $\text{success} = 0.1694 (5.703) + 5.5318E-7 (1.349) \text{ dprev}$. 'Dprev' is statistically not significant at the 5% level, while 'divers' is highly significant if it is the only explaining variable (check B (iii)). Thus, also for the small sample, it can be concluded that diversification has more influence than firm size in explaining the success ratios.

Finally, I have checked whether there is a relation between the success ratio and the profitability of a firm. To correct for the size effect, I have taken as the profitability variable the DP operating income over the period 1986-1990 divided by the DP revenues for that same period. For the sample of all 100 firms, an OLS regression is performed with profitability as dependent and success ratio as independent variable. The coefficient of the success ratio has a negative sign, but is statistically not significant at a 10% level. The adjusted R^2 is very low (about 1%). The same regression is also performed for the smaller sample of 44 firms with 5 and more applications. Again, the coefficient of the success ratio has a negative sign but is now statistically significant at a 5% level. The explained variance of profitability (adj. R^2) is equal to 8.1%. One should, however, be very careful in interpreting these results. It is not necessary so (and even doubtful) that, since the success ratio is inversely related to profitability, pioneering innovation strategies are less profitable than imitative strategies. The country effect can also be important. We saw that US and Japanese firms have lower success ratios in the period 1986-1990. If US and/or Japanese firms are more profitable, we would thus also get a negative sign. For the large sample of 100 firms, I found no evidence for US or Japanese firms being more profitable. For the smaller sample, however, there does seem to be some evidence. US firms are more profitable than European firms (significant at a 5% level), while Japanese firms are not. Thus, there is some evidence that the inverse relationship between success ratio and profitability is related to country differences. In fact, as we will see in the next section, Japanese and US firms have a lower success ratio than European firms over the period 1986-1990 partially because they have to wait longer for their grants.

6.4 National Waiting Times and Success Ratios

The results with respect to the country dummies in the above regressions (the B-series) indicate that US and Japanese data-processing firms have significantly lower success ratios than European firms for their applications during 1986-1990. There are two possible reasons to explain this. First, US and Japanese firms may

have to wait longer for their grants. Fewer of their 1986-1990 applications are then processed, resulting in lower success ratios. This possibility will be examined in subsection (a). A second reason may be that US and Japanese firms just perform worse because of their country of origin, while European firms may perform systematically better in the EPO. This possibility will be examined in subsection (b).

(a) National Waiting Times

The waiting times for the data-processing industry are first determined for the year of application 1986 and are calculated as follows. I have taken all applications of a country in 1986 and checked how many of those were granted a patent in the period 1986 to August 1993. Next, I checked how many of these total grants were granted after how many years. The categories of waiting times are: 2 (and less), 3, 4, 5, 6 and 7 (and more) years. I will examine the waiting times for the US, Japan, Germany, France and the United Kingdom. For each of these countries, I construct a cumulative distribution of their waiting times. Using the Kolmogorov-Smirnov test³, I test in pairs whether these lag distributions differ. The null hypothesis is that distributions do not differ. The alternative hypothesis is that m must wait longer than n , where m and n are the total grants of a country. The Kolmogorov-Smirnov test statistic $D_{m,n}$ is the largest difference between the cumulative distribution of m and that of n . Given the large samples and the one-tailed test I want to perform, I can use the statistic $X^2 = 4D_{m,n}^2 mn/(m+n)$ which is approximated by the Chi-square distribution with 2 degrees of freedom (Siegel and Castellan Jr. 1988).

Table 6.3 summarizes the results. From the Kolmogorov-Smirnov tests in table 6.3, a country ranking can be constructed for the waiting times for the data-processing industry in 1986. In order of decreasing waiting time, the ranking which emerges is: 1. Japan; 2. US; 3. UK; 4. Germany; and 5. France. Only the difference between the US and Japan is not significant at the 5% level. The higher speed of processing European applications, as given by the British, German and French applications, might thus partially explain the higher success ratios of European firms found in the regressions in section 6.3.

(b) National Success Ratios

3. The Kolmogorov-Smirnov test is a non-parametric test which can be used to check whether two (or more) distributions have an identical distribution underlying them. For more details, see Siegel and Castellan Jr. (1988).

Table 6.3

National differences in waiting times in the data-processing industry in 1986

Countries	Kolmogorov-Smirnov Statistic	χ^2	Significance
m = 315 French grants n = 640 German grants	$D_{m,n} = 0.0997$	8.390	$0.02 > p > 0.01$
m = 640 German grants n = 127 UK grants	$D_{m,n} = 0.1268$	6.812	$0.05 > p > 0.02$
m = 127 UK grants n = 950 US grants	$D_{m,n} = 0.1938$	16.837	$p < 0.001$
m = 950 US grants n = 767 Japan. grants	$D_{m,n} = 0.0430$	3.132	$0.30 > p > 0.20$

Table 6.4

National differences in waiting times in the data-processing industry during 1978-1985

Countries	Kolmogorov-Smirnov Statistic	χ^2	Significance
m = 1725 French grants n = 3466 German grants	$D_{m,n} = 0.1167$	62.782	$p < 0.001$
m = 3466 German grants n = 765 UK grants	$D_{m,n} = 0.1545$	59.848	$p < 0.001$
m = 765 UK grants n = 5961 US grants	$D_{m,n} = 0.1153$	36.054	$p < 0.001$
m = 5961 US grants n = 3779 Japan. grants	$D_{m,n} = 0.0762$	53.132	$p < 0.001$

In order to get larger samples and to check whether these differences in processing speed have been existing from the start, I have carried out identical tests for applications filed in the period 1978-1985. The results are presented in table 6.4. Exactly the same order of waiting times also emerges from these samples. From large to small times, the ranking is: 1. Japan; 2. US; 3. UK; 4. Germany; and 5. France. All differences are highly significant ($< 0.1\%$ level).

Is this phenomenon specific to the data-processing industry? Table 6.5 summarizes the results for all applications granted (i.e., in all IPC classes) in the period 1978-1985. Again, the same order as before is found here with very strong evidence. It can thus be concluded that the difference in waiting times is not specific to the data-processing industry.

Table 6.5
National differences in waiting times for all grants during 1978-1985

Countries	Kolmogorov-Smirnov Statistic	χ^2	Significance
m = 14191 French grants n = 33661 German grants	$D_{m,n} = 0.0609$	148.093	$p < 0.001$
m = 33661 German grants n = 9198 UK grants	$D_{m,n} = 0.1636$	773.401	$p < 0.001$
m = 9198 UK grants n = 34785 US grants	$D_{m,n} = 0.1176$	402.416	$p < 0.001$
m = 34785 US grants n = 20660 Japan. grants	$D_{m,n} = 0.0516$	138.045	$p < 0.001$

It should be emphasized here that I have only observed the differences in national waiting times. It is outside the scope of this study whether these differences are due to preferential treatment given by the EPO to European applicants, for example, or to the organizational structure within the EPO, or to any other factors.

(b) National Success Ratios

In order to isolate the differences in EPO success ratios of countries from the differences in waiting times, I will focus on the period 1978-1985. The vast majority of applications during that period were processed by August 1993 for

each country. Thus, differences in waiting times do not affect differences in success ratios. Table 6.6 presents the success ratios of countries in the data-processing IPC classes as well as the overall success ratios of countries.

The overall success ratios do not seem to differ much from the ratios in the data-processing classes for each country. The examination of data-processing files thus appears to be not more or less stringent than that of other files. From table 6.6 it can furthermore be concluded that the lower success ratio of Japan which was found in the data-processing industry in section 6.3 is a severe misrepresentation. Japanese firms generally perform better than European firms, both in data processing and overall.

Table 6.6
EPO success ratios during 1978-1985

Country	Data Processing	Overall
Japan	0.855	0.833
France	0.785	0.762
Germany	0.717	0.731
US	0.659	0.623
UK	0.552	0.581

How can these differences in EPO success ratios of countries be explained? One possible explanation may be found in the novelty requirements of the national patent offices. If applicants are used to stringent examination, they are more likely to carefully select the inventions they file. In a less stringent system, they would then perform better than applicants who are used to looser national examination. For countries outside the EPC, one might bring forward the argument that their national offices have already screened the applications, because those applicants may be expected to first file in their national office and then in the EPO.⁴ The better the screening (i.e., the higher the novelty requirements), the higher the success ratio for that country in the EPO. I thus expect lower national success ratios to lead to higher EPO success ratios. The national success ratios are taken from the WIPO publication "100 Years Protec-

4. This is also common practice for much European applicants.

tion of Industrial Property" (Geneva, 1983), which contains the total number of applications and patents granted in a country for the period 1883-1982. I assume that the national patent offices treated residents and non-residents in the same way. The national success ratios then indicate the experience of national applicants with their patent office. The EPO success ratios have been calculated for the period 1978-1985 because, as indicated previously, almost all applications from that period were processed by August 1993. The sample contains 36 countries (all countries I could find data for, both in the EPO and in the WIPO publications).⁵ For reasons outlined above, a two-limit Tobit model is used again. With the EPO success ratios ('eposuccess') as dependent and the national success ratios ('natsuccess') as independent variable, the estimation is given by:

$$\text{eposuccess} = 0.7392 \text{ (10.891)} - 0.2581 \text{ (-2.648)} \text{ natsuccess}$$

The t-ratios are given in parentheses. The coefficient of 'natsuccess' has the expected sign and is statistically significant at a 0.5% level. It can thus be concluded that national success ratios are indeed of influence in explaining the EPO success ratios.

The argument brought forward above might be called 'Porterian' in the sense that stronger domestic screening implies better foreign performance. Another, more indirect and speculative, reason behind the inverse relationship might be found in the type of incentive provided by a patent system. The choice between the two innovation strategies, pioneering or imitative, may depend on the dimensions of the patent system, as explained in chapter 5. If a patent system sets high barriers for applicants, by means of high novelty requirements, the pioneering strategy might become more costly, but provides more protection against further improvement in return. Firms that pursue pure pioneering strategies benefit from the advantage of high protection without investing more in basic research (because they will meet the novelty requirements anyway). Therefore, stringent novelty requirements can be thought to encourage pioneering strategies. Indirectly, strong domestic novelty requirements then lead to better patenting performance in the EPO, by inducing pioneering innovation strategies.

5. The following countries have been included: Argentina, Australia, Austria, Belgium, Brasil, Bulgary, Canada, Cyprus, (former) Czechoslovakia, Denmark, Finland, France, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Luxemburg, Malta, Netherlands, New Zealand, Norway, Poland, Portugal, (former) Soviet Union, South Africa, Spain, Sweden, Switzerland, Turkey, UK, US, (former) West-Germany, (former) Yugoslavia.

6.5 Conclusions

This chapter has argued that firms which pursue pioneering innovation strategies are more successful in obtaining patent protection than firms which pursue imitative innovation strategies. A pioneering strategy aims at profits resulting from new or improved products and processes. This strategy requires more basic research, which explores novel and unknown paths that are only weakly related to existing products, and therefore its fruits are better patentable. The novelty requirements are met more easily. An imitative strategy, on the other hand, aims at modifying and redesigning existing products and makes more use of applied research and development. Due to the stronger connection with existing products, the fruits from applied research and development pass the patent office less easily.

This theory has been tested for the data-processing industry in the period 1986-1990. Since no direct data on basic vs. applied research were available, the success ratios of data-processing firms in the EPO are explained by indirect variables which indicate the degree of basic research. One such variable is diversification. As argued in Nelson (1959) and Arrow (1962), more diversified firms are expected to do more basic research because they can better spread the risks and apply the results in more fields. Indeed, the variable for diversification within the data-processing market turns out to be the most important in explaining the success ratios. Another variable, the propensity to patent, also performs well. Pioneering research does not only result in more patent applications but also in better patenting performance.

Another result is that US and Japanese firms achieved lower success ratios for their applications in the period 1986-1990. However, it is found that this is due to the longer waiting times between applications and patents granted in the EPO for US and Japanese firms, compared to European firms. Without this difference in waiting times, Japanese applicants in general have higher success ratios than European and US applicants. Furthermore I have found some evidence that national patent offices act as screening institutes for the EPO. Countries without stringent screening, visualized by low novelty requirements, perform worse in the EPO than countries that apply more stringent screening.

The question may be raised why countries set different patent heights. In the conclusion of chapter 3, I suggested that countries that perform relatively well in basic research might set high protection, whereas countries that are more specialized in applied research and development might set low protection. In the next chapter, a model is developed that examines international differences in patent protection. Although the analysis focuses on patent breadth, the conclusions also hold for other dimensions, such as length and height.

An Address Model of Patent Breadth with Two Countries

Drawing on Hotelling's model of spatial competition, we develop an address model in which normative and positive aspects of the choice of patent breadth in an international setting can be examined. In each of two countries a number of different entrepreneurs face the following sort of opportunity: by investing a sum c in product research and development the entrepreneur makes possible the production of an entire spectrum of products. Any entrepreneur's incentive to do so is determined jointly by patent breadths in the two countries where the spectrum of products will be sold. As before, patent breadth is measured by the portion of the product spectrum over which the patentholder exercises monopoly power. In each country there is a population of entrepreneurs, each of whom can develop a different product spectrum; entrepreneurs differ with regard to the magnitude of the development cost c that must be incurred to create the spectrum of products.

Even in a one-country setting this poses a difficult and typical policy problem because there are many incentive problems to be solved, one for each entrepreneur, while there are two policy instruments, patent breadths in the two countries. To complicate matters, if we assume that in making decisions with regard to patent breadth, each country has regard only to the welfare of its own citizens, significant international externalities distort the policy-setting game.

Given patent breadths in the two countries, only those entrepreneurs who (correctly) anticipate that they can cover their development costs do, in fact, develop (and, of course, patent) their separate product spectrums. Knowing the distribution of opportunities and associated development costs in the two countries, and knowing how entrepreneurs will respond to the incentives created by patent breadth, each country non-cooperatively chooses its own patent breadth to maximize a weighted sum of profit and consumers' surplus earned by the country's citizens.

We establish the following results. First, when countries place equal weight

on profit and consumers' surplus of their own citizens, relative to the social optimum, patent breadths are too narrow. This result reflects the existence of a positive externality flowing from each country's patent breadth to the profit and consumers' surplus enjoyed by citizens of the other country. Second, the first result generalizes to other facets of intellectual property rights. Intellectual property rights tend to be too weak if countries non-cooperatively choose them to maximize total surplus of their own citizens. Third, to the extent that profit is weighted more heavily than consumers' surplus in political objective functions, the tendency for property rights to be too weak in equilibrium is offset. Indeed, if the weight given to profit in the political objective function is large enough, property rights will be too strong in equilibrium. The final result is that, except in very special circumstances, equilibrium patent breadths are not identical in the two countries, and in all cases there is a continuum of optimal mixes of patent breadth, only one of which is symmetric. Further, if equilibrium patent breadths are sufficiently asymmetric, there is no symmetric patent policy that Pareto-dominates the original equilibrium.

This chapter is structured as follows. Section 7.1 discusses the literature that is related to the questions posed here. The core model with one country is presented in section 7.2. Section 7.3 examines the entry game, i.e., the stage where firms make decisions on their innovations. The stage where two countries choose their patent policies interdependently is studied in section 7.4. Some concluding remarks and possible applications are discussed in the final section 7.5.

7.1 Related Literature

Chapter 2 has already discussed the theoretical patent models from the Industrial Organization literature. All these models (for example, Nordhaus 1969, Klemperer 1990, Scotchmer and Green 1990), essentially analyse a trade-off between innovation incentive and monopolistic welfare loss. In an international context, this trade-off between R&D incentives and deadweight losses occurs at the level of global welfare, but is complicated by the distribution of benefits and costs over individual countries. Extending Nordhaus (1969), Berkowitz and Kotowitz (1982) have examined the optimal patent length for a small country. The size of process innovation is a function of patent lengths in a small country and in the rest of the world. The small country assumption ensures that no retaliation of other countries is to be expected if the country shortens the national patent life. Whether a single country should set long or short patent lifetimes depends on the industrial organization of the domestic invention market. Berkowitz and Kotowitz conclude that a competitive invention industry calls for short-lived patents and a concentrated invention industry for long-lived patents.

Inspired by the contentious discussion on intellectual property rights in the Uruguay Round, Chin and Grossman (1990), Diwan and Rodrik (1991) and Dear-dorff (1992) have developed models which analyse patent protection in north-south trade relations. Many southern countries do not provide patent protection; they either have no patent system or do not enforce it effectively. Their claim is that southern welfare only decreases if patent protection is provided for northern innovations that will be generated anyway. Northern countries, particularly those representing northern firms, want protection against imitation in the south in order to be able to recoup their R&D investments.

In the present chapter patent protection is defined within a model of horizontal product differentiation similar to Klemperer's (1990) model. Both types of welfare loss in his analysis, i.e., substitution to less preferred varieties and to not buying, can potentially arise in our model. Unlike Berkowitz and Kotowitz (1982) we do not use a small country assumption. Hence, the strategic interactions between two countries which determine their patent breadths are taken fully into account. For certain parameter configurations, for example, when demand in one country is very weak, our model comes close to the setting studied by Berkowitz and Kotowitz. Our model can also be used to investigate the optimal protection levels for the north and the south. Unlike previous studies in that area, we do not assume beforehand that only the north has potential innovators. This situation occurs in our model as a special case, more precisely when the innovation density in the south is zero, as will become clear below.

7.2 The Core Model

In this section we focus on just one country as we develop the core of our model. In passing, we comment briefly on the choice of patent breadth in a one-country setting. By spending c on product development the entrepreneur can create an entire spectrum of products. An individual product is described by its address w in the product spectrum, which extends from $-\sigma$ to $+\sigma$, where parameter $\sigma > 0$.

If the product is developed, we assume that any firm is technically able to produce any product in the spectrum at a marginal cost, which we set equal to 0 for convenience. Hence, in the absence of some form of property right, the entrepreneur has no incentive to develop the spectrum: with no property right to any portion of the product spectrum, the equilibrium price for all products in the spectrum is 0. Similar to the analysis in chapter 4, we suppose such property rights take the form of a monopoly right granted to the developer to produce products in interval $[-b/2, b/2]$, where $0 < b < \sigma$. Thus, subsequent to development, the entrepreneur has the right to monopolistically exploit a market of length b .

As in Hotelling's (1929) model, there is a continuum of consumers. Each consumer has a most preferred product in each spectrum, which we denote by w^* . The demand side of the model is created by positing a continuum of consumers with a uniform distribution of w^* on $[-\sigma, \sigma]$. We set the density of this distribution equal to 1. Denote the price of product with address w , or more simply product w , by $p(w)$ (for w in $[-\sigma, \sigma]$). Given prices, the consumer first chooses which product to buy, and then chooses how much to buy of the product chosen in stage one. Consumer w^* buys the good w for which

$$p_d(w, w^*) = p(w) + t|w - w^*| \quad (7.1)$$

is smallest. The parameter t is analogous to the cost of transport in familiar models of spatial competition; here, of course, it is a utility penalty associated with consuming a product that differs from the consumer's most preferred product. And the function $p_d(w, w^*)$ is analogous to a delivery price schedule.

If for two goods, w^1 and w^2 ,

$$p(w^1) + t|w^1 - w^*| = p(w^2) + t|w^2 - w^*| \quad (7.2)$$

we assume that consumer w^* prefers the good that is nearer to his most preferred good, while if they are equidistant from w^* the consumer randomly picks one of them.

The demand function for the consumer's *chosen* product is the following linear function of p_d :

$$x(p_d) = (a - p_d)/F(n) \quad (7.3)$$

Parameter a is the intercept of the inverse demand function and $F(n)$ is its slope. The slope of the inverse demand function, $F(n)$, is an increasing function of n , the number of developed spectra:

$$F(n) = \gamma n^\theta \quad 0 < \theta < 1 \quad \gamma > 0 \quad (7.4)$$

Notice that as the number of developed spectra increases, the consumer buys less of his chosen product in each spectrum. The larger parameter θ , the more pronounced this demand interdependence effect is. This inter-spectrum demand effect assures that the marginal benefit of an additional product spectrum decreases as the number of developed spectra increases.

Given patent breadth b , equilibrium prices are 0 in the two unprotected intervals of the product spectrum, intervals $[-\sigma, -b/2]$ and $[b/2, \sigma]$, since any firm

can produce these goods at 0 marginal cost. In the protected interval the monopolist maximizes profit by providing all conceivable varieties and by choosing $p(w)$ equal to $a/2$ if the consumer with $w^* = w$ prefers buying the monopolist's good at this price rather than buying good $w = b/2$ (or good $w = -b/2$) at price 0. If, however, the monopolist is constrained by the possibility that the customer with $w^* = w$ would prefer the nearest competitively priced good to the monopolist's good at price $a/2$, to maximize profit the monopolist sets $p(w)$ equal to $t|w - b/2|$ (or $t|w + b/2|$). Thus, equilibrium prices in the monopolist's protected segment are

$$p(w) = \min [a/2, t|w - b/2|, t|w + b/2|]$$
 (7.5)

The first term is the unconstrained monopoly price, and the second (respectively, third) is the highest price which will induce the consumer with address $w^* = w$ to choose the monopolist's good w in preference to good $b/2$ (respectively, $-b/2$), which is available at price 0). In this price equilibrium, each consumer buys her most preferred good. The price equilibrium is illustrated in figure 7.1.a, for the case in which $tb > a$, and in figure 7.1.b, for the case in which $tb < a$.

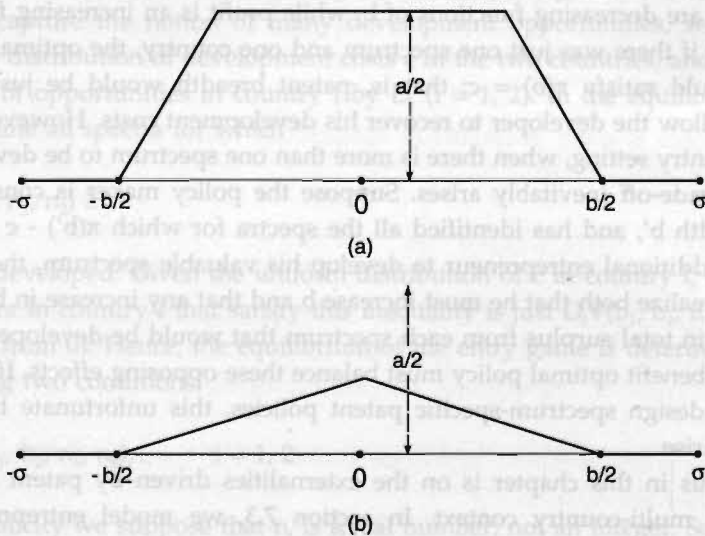


Figure 7.1 Equilibrium price schedules

At this point we have seven parameters: a , t , θ , b , γ , σ and c . The preference parameters θ and γ in equation (7.4) are common to both countries throughout the analysis, while, in the following sections of the chapter, the preference parameters a and σ and the policy variable b are country-specific. Experience with the model suggests that very little is lost if we choose parameter t to be infinite. In this case, the monopolist's profit-maximizing price is $a/2$ throughout the protected portion of the spectrum, which permits a significant simplification of the algebra in the balance of the analysis. Accordingly, we set t equal to infinity. Given this simplification, the developer's gross profit from a country with patent breadth b is just

$$\pi(b) = (a^2 b) / (4F(n)) \quad (7.6)$$

Consumers' surplus from the protected, monopoly portion of the spectrum is $(a^2 b) / (8F(n))$, and from the unprotected, competitive portion of the spectrum it is $(a^2 (2\sigma - b)) / (2F(n))$. Hence, the consumers' surplus from a product spectrum breadth b is

$$\phi(b) = (a^2 / 8F(n)) (8\sigma - 3b) \quad (7.7)$$

Notice that total surplus (the sum of profit and consumers' surplus) and consumers' surplus are decreasing functions of b , while profit is an increasing function of b . Hence, if there was just one spectrum and one country, the optimal patent breadth would satisfy $\pi(b) = c$; that is, patent breadth would be just broad enough to allow the developer to recover his development costs. However, even in a one-country setting, when there is more than one spectrum to be developed, a difficult trade-off inevitably arises. Suppose the policy maker is considering patent breadth b' , and has identified all the spectra for which $\pi(b') - c > 0$. To induce an additional entrepreneur to develop his valuable spectrum, the policy maker will realize both that he must increase b and that any increase in b entails a reduction in total surplus from each spectrum that would be developed given b' . The cost-benefit optimal policy must balance these opposing effects. If it were possible to design spectrum-specific patent policies, this unfortunate trade-off would not arise.

Our focus in this chapter is on the externalities driven by patent breadth policy in a multi-country context. In section 7.3, we model entrepreneurial decisions given the opportunity to sell in different countries with (possibly) different patent policies. Following that, in section 7.4 we examine the two-country policy game in which countries choose patent policies.

7.3 The Entry Game

Here we suppose there are two countries, 1 and 2, and look at the game in which the number of spectra that entrepreneurs choose to develop in the two countries is determined. The two parameters in equation (7.4), θ and γ , are assumed to be common to the two countries, while the remaining parameters are country-specific: a_1 and a_2 are intercepts of the inverse demand functions; σ_1 and σ_2 are spectrum length parameters. In the entry game of this section, patent breadths in the two countries, b_1 and b_2 , are exogenous. In the following section we build on the results of the entry game to determine equilibrium values of these policy parameters.

We assume that the developer of any spectrum will receive local patent protection in both countries. It is this assumption that drives the inter-country externalities in our model. An immediate consequence of the assumption is that any spectrum that is developed will be available in both countries. Hence, from equation (7.6), the developer's aggregate gross profit or inducement to innovate, which we denote by $V(b_1, b_2; n_1, n_2)$, is given by the following expression

$$V(b_1, b_2; n_1, n_2) = (b_1(a_1)^2 + b_2(a_2)^2)/(4F(n_1 + n_2)) \quad (7.8)$$

where n_i ($i = 1, 2$) is the number of spectra developed by citizens of country i .

To capture the notion of many development opportunities, we assume a uniform distribution of development costs c in the two countries, and denote the density of opportunities in country i by D_i ($i = 1, 2$). In the equilibrium of the entry game all spectra for which

$$V(b_1, b_2; n_1, n_2) \geq c$$

will be developed. Given the uniform distribution of c in country i , the number of spectra in country i that satisfy this inequality is just $D_i V(b_1, b_2; n_1, n_2)$ (when starting from 0). Hence, the equilibrium of the entry game is determined by the following two conditions:

$$n_i = V(b_1, b_2; n_1, n_2), \quad i = 1, 2 \quad (7.9)$$

For simplicity we suppose that n_i is a real number, not an integer. Solving these entry equilibrium conditions, and letting N_i denote the equilibrium value of n_i , we get

$$N_i = (D_i(Z)^{1/(1+\theta)})/(D_1 + D_2)^{\theta/(1+\theta)} \quad (7.10.a)$$

where

$$Z = (b_1(a_1)^2 + b_2(a_2)^2)/(4\gamma) \quad (7.10.b)$$

From (7.10) we get a number of useful comparative static results:

(i) N_i increases with both policy parameters, b_1 and b_2 . Further, the rate at which N_i increases as b_k ($k = 1, 2$) increases is proportional to $(a_k)^2$; this reflects the facts that developers who reside in country i have access to patent protection in both countries, and that protection is more valuable as the individual consumers' demand for the product is higher. Conversely, N_i increases with a_k ($k = 1, 2$) and the rate of increase is proportional to b_k . Finally, N_i is concave in b_k ($k = 1, 2$).

(ii) Naturally, N_i increases as D_i increases. And N_i decreases as D_k (k not equal to i) increases. This latter result is driven by inter-spectrum demand dependence; as D_k increases, so does N_k , but a larger N_k shifts individual demand down in both countries, thereby decreasing N_i .

(iii) Given that b_k is less than $2\sigma_k$ (spectrum length), N_i is independent of σ_k . Given that the protected portion of the spectrum is smaller than the entire spectrum, the size of the spectrum has no bearing on profitability and hence no bearing on the number of developers that find it profitable to develop their spectra.

The inducement to enter evaluated at equilibrium values for N_1 and N_2 , which we denote by $V^*(b_1, b_2)$, is

$$V^*(b_1, b_2) = (Z^{1/(1+\theta)})/(D_1 + D_2)^{\theta/(1+\theta)} \quad (7.11)$$

This is the gross profit of each of the developed spectra in equilibrium. Of course, net profit differs from spectrum to spectrum due to differences in development costs. At one extreme, net profit is identical to $V^*(b_1, b_2)$, and at the other it is 0.

We can readily compute aggregate development costs in equilibrium for country i . They are equal to

$$C_i = (D_i(Z)^{2/(1+\theta)})/(2(D_1 + D_2)^{2\theta/(1+\theta)}) \quad (7.12)$$

The total net profit accruing to developers in country i , $\pi_i(b_1, b_2)$, is then $N_i^*(b_1, b_2) - C_i$, which reduces to

$$\pi_i = (D_i(Z)^{2/(1+\theta)}) / (2(D_1 + D_2)^{2\theta/(1+\theta)}) \quad (7.13)$$

The total net profit for country i inherits the comparative static properties of N_i : both N_i and π_i are increasing in b_i , a_i , and D_i , decreasing in D_k (k not equal to i), and independent of σ_i . Interestingly, despite the inter-spectrum demand dependence and the ever-increasing marginal cost of an additional developed spectrum, $\pi_i(b_1, b_2)$ is convex in b_k ($k = 1, 2$) (recall that $\theta < 1$). That is, aggregate profit of country i increases at an increasing rate as patent breadth in either country increases.

Finally, since all developed spectra are available in both countries, from (7.7) we see that aggregate consumers' surplus for country i is

$$\phi_i(b_1, b_2) = (N_1 + N_2)((a_i)^2 / (8F(N_1 + N_2)))(8\sigma_i - 3b_i) \quad (7.14)$$

Using previous results this can be rewritten as

$$\phi_i(b_1, b_2) = W(a_i)^2(Z(D_1 + D_2))^{(1-\theta)/(1+\theta)} \quad (7.15.a)$$

where

$$W = (8\sigma_i - 3b_i) / 8\gamma \quad (7.15.b)$$

The consumers' surplus comparative statics are also interesting:

(i) Consumers' surplus of country i is increasing in b_k (k not equal to i). An increase in country k 's patent breadth increases the number of developed spectra available in country i , while having no direct effect on consumers' surplus per spectrum in country i , thus increasing consumers' surplus in country i . (There is, of course, an indirect effect on consumer surplus per spectrum driven by the downward shift of the individual demand curve as the number of spectra increases, but the increase in consumers' surplus resulting from the increase in the number of developed spectra dominates this indirect effect.)

(ii) Provided that b_k is not too large, consumers' surplus in country i (i not equal to k) initially increases as b_i increases and subsequently decreases. Here there are two offsetting effects: consumers' surplus in country i tends to increase as b_i increases because the increase in patent breadth results in an increase in the number of developed spectra in country i ; offsetting this effect is the fact that surplus per available spectrum in country i decreases as b_i increases. The second effect eventually dominates the first, and if b_k is sufficiently large, it dominates

from the outset.

(iii) Naturally, consumers' surplus in country i increases as a_i increases. In addition, it increases as a_k (k not equal to i) increases. The causality runs through the effect of a_k on N_i ($i = 1, 2$): an increase in a_k results in an increase in the number of spectra developed in both countries, thereby increasing consumers' surplus in country i .

(iv) Understandably, consumers' surplus increases in country i increases as σ_i , the width of the product spectrum in country i , increases, and is independent of σ_k (k not equal to i).

In the next section we build on these results as we analyze the policy game in which patent breadths are determined.

7.4 The Policy Game

In this section we analyze the game in which patent breadths are chosen independently and non-cooperatively by the two countries to maximize a weighted sum of profit and consumers' surplus of the country's citizens. Letting λ_1 and λ_2 denote the weights given to profit in the political objective functions in the two countries, we get the following objective functions:

$$W_i(b_1, b_2) = \lambda_i \pi_i(b_1, b_2) + (1 - \lambda_i) \phi_i(b_1, b_2) \quad i = 1, 2 \quad (7.16)$$

For purposes of comparison we also need the total surplus functions for the two countries:

$$\omega_i(b_1, b_2) = \pi_i(b_1, b_2) + \phi_i(b_1, b_2) \quad i = 1, 2 \quad (7.17)$$

Of course, when equal weight is given to profit and consumers' surplus (when λ_i is equal to $1/2$), the objective of the policy maker is the maximization of total welfare.

Profit and Consumers' Surplus Weighted Equally

We begin by looking at the case in which profit and consumers' surplus are weighted equally in the political objective functions. The important qualitative features of the equilibrium of this game are illustrated in figure 7.2. In this equilibrium, country 1 chooses $b_1 = 8.9$ and country 2 chooses $b_2 = 5.2$. Although the parameter values on which this diagram is based are reported in the figure,

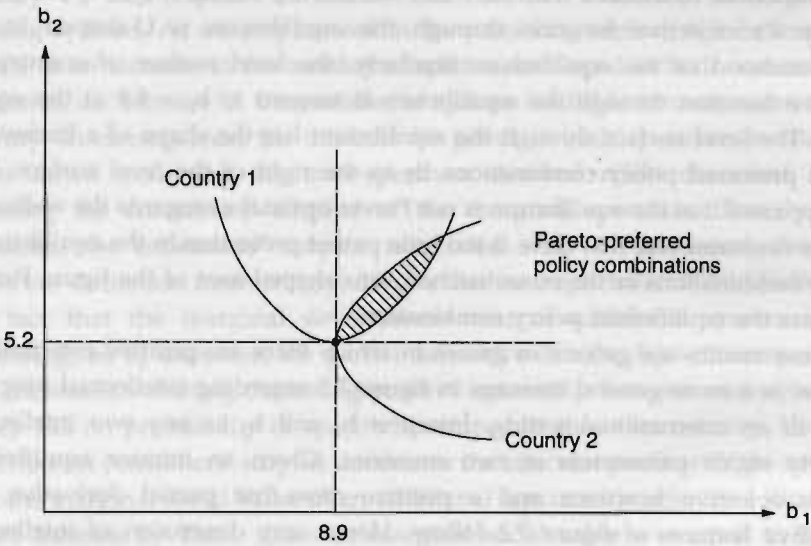


Figure 7.2 Asymmetric innovation densities ($\gamma = 10$; $\theta = 0.5$; $a_1 = a_2 = 10$; $\sigma_1 = \sigma_2 = 30$; $\lambda_1 = \lambda_2 = 0.5$; $D_1 = 1.5$; $D_2 = 1.0$)

we will not focus on them since, at this point, we are interested in the qualitative properties of equilibrium.

One useful way to regard this equilibrium is to think of each country as maximizing its own objective function, subject to the constraint that the other chooses its equilibrium value of b . Focusing for concreteness on country 1, it is then apparent that a level surface (or indifferent curve) of country 1's objective function is tangent to the constraint $b_2 = 5.2$ at the equilibrium. Importantly, the curvature of country 1's level surface in the neighbourhood of the equilibrium, and the set of policy combinations that are preferred by country 1 to the equilibrium policy combination, are determined by the sign of the partial derivative of $W_1(b_1, b_2)$ with respect to b_2 . The sign of this derivative is, of course, positive since both profit and consumers' surplus for country 1 increases as b_2 increases. For country 1 then, the set of preferred policy combinations lies above 1's level surface through the equilibrium. Now, beginning at the equilibrium suppose we alter (increase or decrease) b_1 slightly: $W_1(b_1, b_2)$ necessarily decreases since, given $b_2 = 5.2$, $b_1 = 8.9$ maximizes $W_1(b_1, b_2)$; then, to get back to the original level surface, the positive sign of the partial derivative of $W_1(b_1, b_2)$

with respect to b_2 dictates that we must increase b_2 ; hence, the level surface of country 1's objective function through the equilibrium is U-shaped in the neighbourhood of the equilibrium. Similarly, the level surface of country 2's objective function through the equilibrium is tangent to $b_1 = 8.9$ at the equilibrium. The level surface through the equilibrium has the shape of a backwards C, and preferred policy combinations lie to the right of the level surface. It is then apparent that the equilibrium is not Pareto-optimal as regards the welfare of the two countries and that there is too little patent protection in the equilibrium - policy combinations in the cross-hatched, lens-shaped area of the figure Pareto-dominate the equilibrium policy combination.

These results are generic in games in which there are positive externalities. So there is a more general message in figure 7.2 regarding intellectual property rights in an international setting. Interpret b_1 and b_2 as any two intellectual property rights parameters in two countries. Given an interior equilibrium, smooth objective functions and a positive cross-first partial derivative, the qualitative features of figure 7.2 follow. Hence, any dimension of intellectual property rights that confers a positive externality on the welfare of other countries - length of patent, for example - will tend to be too weak in equilibrium, if countries choose such policies non-cooperatively to maximize aggregate welfare of their own citizens. As we argue below, this conclusion does not necessarily follow if profit is weighted more heavily than consumers' surplus in the political objective function.

The comparative statics of the policy equilibrium are also of some interest. Our approach in figures 7.3 through 7.6 is to use the equilibrium of what we call the baseline model as a point against which to compare the equilibria that emerge as we change various parameters of the model. Although we used numerical techniques to generate results, they are readily understood. In figure 7.3 we have drawn the best-response functions for the baseline model, the solid lines in the figure that generate the equilibrium E_0 . We have also drawn the best-response functions that arise when we increase spectrum width in country 1, the dashed line for country 1 and again the solid line for country 2 that generate the equilibrium at E_1 . An increase in spectrum width in country 1 leads to an increase in the equilibrium value of b_1 and a decrease in the equilibrium value of b_2 .

The first thing to notice about figure 7.3 is that the best-response functions are downward sloping. This reflects the fact that an increase in one country's policy parameter not only increases welfare in the other country, but also decreases the rate at which the other country's welfare changes as it increases its own policy parameter. It is this second effect that results in downward-sloping best-response functions, and it is readily understood. Suppose we are initially on

country 1's best-response functions, and consider a discrete increase in b_2 , holding b_1 fixed. As we emphasized in the preceding section, this increase in b_2 results in an increase in the number of spectra developed in both countries and, of course, an increase in the number of spectra available to consumers in each country. In consequence, the demand for goods in each spectrum in country 1 shifts down (reflecting the inter-spectrum demand effect discussed above), reducing the consumers' surplus of the marginal spectrum. That is, as b_2 increases, the consumers' surplus associated with an additional developed spectrum decreases. Interestingly, despite this inter-spectrum demand effect and the fact that the marginal development cost of additional spectra is ever increasing, as b_2 increases, the partial derivative of $\pi_i(b_1, b_2)$ with respect to b_1 also increases, reflecting the strong positive externality that one country's patent policy has on profitability of spectra developed in the other country. However, the effect on consumers' surplus dominates the effect on profit with the net result that the marginal social value of an additional spectrum in country 1 diminishes as we increase b_2 . Then, since the marginal social value was equal to 0 in the initial position on country 1's best response, country 1's best response to this exogenous increase in b_2 is to decrease b_1 . That is, country 1's best-response function is downward sloping.

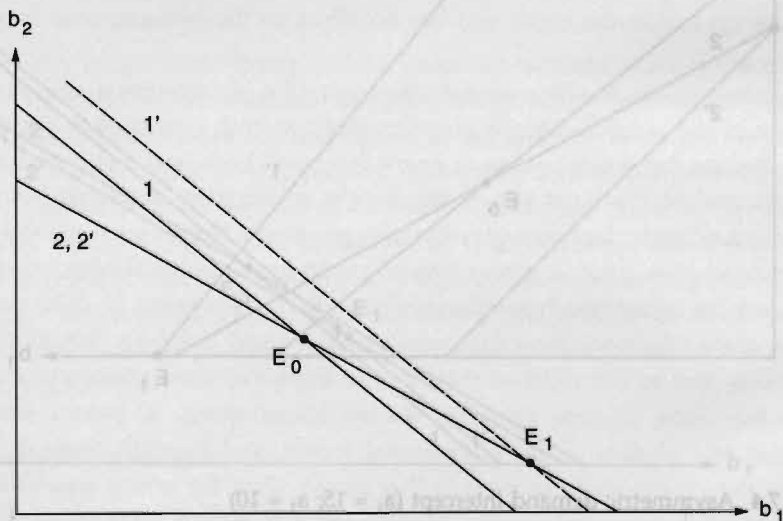


Figure 7.3 Asymmetric spectrum width ($\gamma = 10$; $\theta = 0.5$; $a_1 = a_2 = 10$; $\lambda_1 = \lambda_2 = 0.5$; $D_1 = D_2 = 1.0$; $\sigma_1 = \sigma_2 = 30$ (solid lines: baseline model); $\sigma_1 = 40$; $\sigma_2 = 30$ (dashed lines: asymmetric spectrum width))

From figure 7.3 we see that the increase in σ_1 shifts country 1's best-response function to the right, while having no effect on country 2's best-response function. The key to understanding these effects is that while σ_1 clearly increases consumers' surplus per spectrum in country 1, it has no effect on the inducement to entry (or gross profit) and therefore none on the number of developed spectra. The increase in consumers' surplus per spectrum in country 1 translates into an increase in the marginal social value of an additional spectrum in that country, and hence into a rightward shift in that country's best-response function since b_1 is the country's only control on development. Further, since all the externalities from one country to another are channelled through the number of developed spectra and since σ_1 has no effect on the number of developed spectra, country 2's objective function and therefore its best-response function are independent of σ_1 .

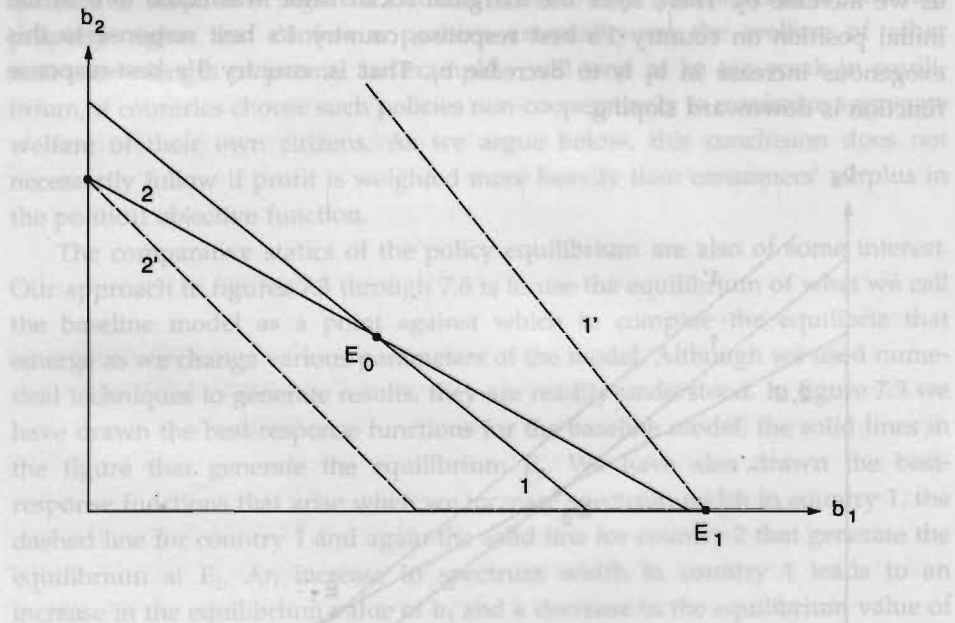


Figure 7.4 Asymmetric demand intercept ($a_1 = 15$; $a_2 = 10$)

In figure 7.4 the solid lines are the best-response functions for the baseline model and the dashed lines the best-response functions that arise when the intercept of the inverse demand function in country 1, a_1 , increases. As we saw in the previous section, this increase in a_1 results in an increase in both N_1 and N_2

(given b_1 and b_2). In country 2 this results in a decrease in the social value of an additional spectrum, and country 2's best-response function shifts downward. In country 1, the increase in a_1 increases the social value of an additional spectrum and the best-response function shifts rightward. Both shifts work to produce an asymmetric equilibrium. In the case illustrated the asymmetry is extreme - b_2 is equal to zero.

As illustrated in figure 7.5, an increase in innovation density in country 1, D_1 , has a similar effect on the equilibrium. Country 2's best-response function shifts downward and 1's shifts rightward producing an asymmetric equilibrium in which b_1 increases and b_2 decreases.

It is then apparent that, except when both countries are identical, the equilibrium of the policy game is asymmetric. Further, if the equilibrium is sufficiently asymmetric there will be no symmetric policy combination that Pareto-dominates the initial equilibrium. This suggests that it will be difficult and perhaps unwise to attempt to create uniform policies with respect to intellectual property rights across very different countries.

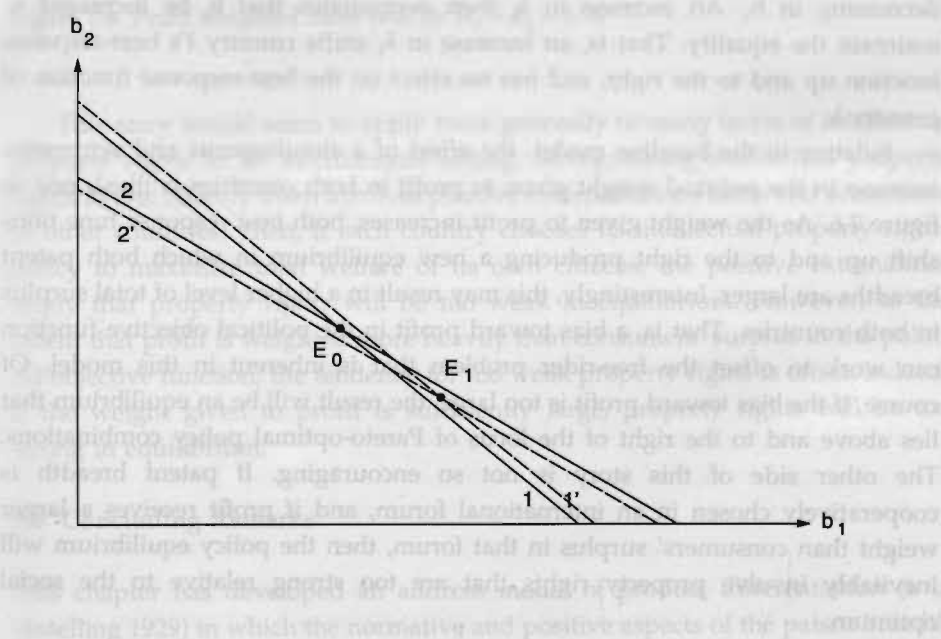


Figure 7.5 Asymmetric innovation densities ($D_1 = 1.5$; $D_2 = 1.0$)

Profits Weighted More Heavily Than Consumers' Surplus

One crude way to introduce the possibility that strongly focused interests receive more weight than diffuse interests in the political objective function is to suppose that countries choose policies to maximize a weighted sum of profit and consumers' surplus. In the current context it seems sensible to suppose that profit receives the larger weight; that is, that λ_i ($i = 1, 2$) exceeds $1/2$. In this subsection we explore some implications of profit-biased weights in the political objective functions.

Notice first that $\pi_i(b_1, b_2)$ and $\phi_i(b_1, b_2)$ are independent of both profit weights (λ_1 and λ_2). Hence, changes in profit weights have no direct impact on the endogenous variables that are determined in the entry game analyzed in section 7.3. The real effects of changes in the weight given to profit in the political objective functions work entirely through their impact on equilibrium policy parameters.

Given b_k (k not equal to i), on country i 's best-response function b_i is such that the partial derivative of $\pi_i(b_1, b_2)$ with respect to b_i , weighted by λ_i , is equal to the partial derivative of $\phi_i(b_1, b_2)$, weighted by $(1 - \lambda_i)$. Necessarily, then, on i 's best-response function profit is increasing in b_i and consumers' surplus is decreasing in b_i . An increase in λ_i then necessitates that b_i be increased to maintain the equality. That is, an increase in λ_i shifts country i 's best-response function up and to the right, and has no effect on the best-response function of country k .

Relative to the baseline model, the effect of a simultaneous and symmetric increase in the political weight given to profit in both countries is illustrated in figure 7.6. As the weight given to profit increases, both best response functions shift up and to the right producing a new equilibrium in which both patent breadths are larger. Interestingly, this may result in a higher level of total surplus in both countries. That is, a bias toward profit in the political objective function can work to offset the free-rider problem that is inherent in this model. Of course, if the bias toward profit is too large, the result will be an equilibrium that lies above and to the right of the locus of Pareto-optimal policy combinations. The other side of this story is not so encouraging. If patent breadth is cooperatively chosen in an international forum, and if profit receives a larger weight than consumers' surplus in that forum, then the policy equilibrium will inevitably involve property rights that are too strong relative to the social optimum.

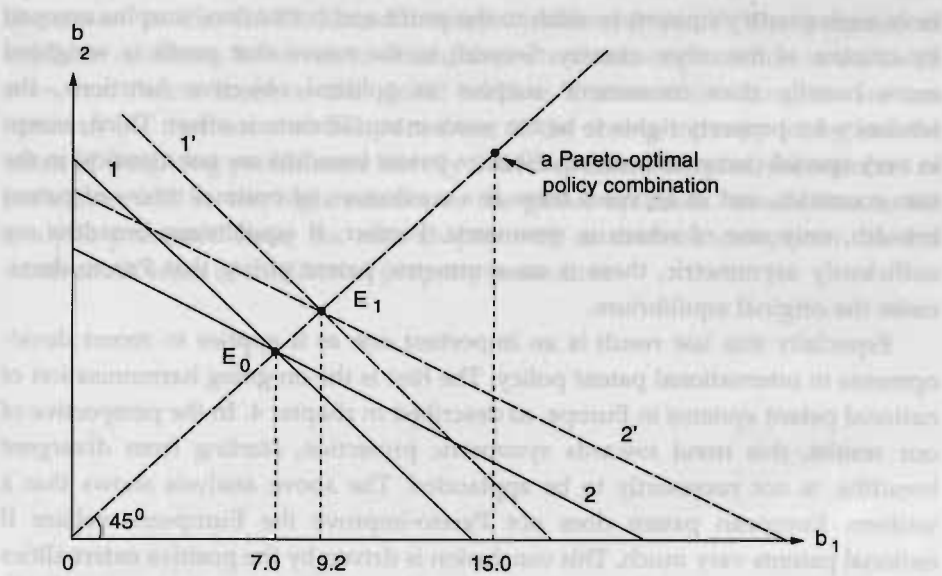


Figure 7.6 Profit weighted more heavily ($\lambda_1 = \lambda_2 = 2/3$)

This story would seem to apply more generally to many facets of intellectual property rights in an international setting. Strengthening intellectual property rights in one country often involves positive externalities on firms and consumers in other countries. Then, if each country chooses its intellectual property rights policy to maximize total welfare of its own citizens, the positive externalities imply that property rights will be too weak in equilibrium. However, to the extent that profit is weighted more heavily than consumers' surplus in the political objective function, the tendency for too weak property rights is offset. Indeed, if the weight given to profit is sufficiently large, property rights will be too strong in equilibrium.

7.5 Concluding Remarks

This chapter has developed an address model of product differentiation (à la Hotelling 1929) in which the normative and positive aspects of the patent breadth policies for two countries can be studied. We have established the following results. First, when countries place equal weight on profit and consumers' surplus of their own citizens, relative to the social optimum, patent breadths are too narrow. This result reflects the existence of a positive externality flowing

from each country's patent breadth to the profit and consumers' surplus enjoyed by citizens of the other country. Second, to the extent that profit is weighted more heavily than consumers' surplus in political objective functions, the tendency for property rights to be too weak in equilibrium is offset. Third, except in very special circumstances, equilibrium patent breadths are not identical in the two countries, and in all cases there is a continuum of optimal mixes of patent breadth, only one of which is symmetric. Further, if equilibrium breadths are sufficiently asymmetric, there is no symmetric patent policy that Pareto-dominates the original equilibrium.

Especially this last result is an important one as it applies to recent developments in international patent policy. The first is the on-going harmonisation of national patent systems in Europe, as described in chapter 4. In the perspective of our results, this trend towards symmetric protection, starting from divergent breadths, is not necessarily to be applauded. The above analysis shows that a uniform European patent does not Pareto-improve the European welfare if national patents vary much. This conclusion is driven by the positive externalities that arise in non-cooperative policy games played by independent countries. In addition to the welfare loss due to harmonisation, another welfare loss may occur if the European Community acts as a forum for producers. Relative to the social optimum, too strong patent protection may emerge if profits are weighted more heavily than consumers' surplus in the total surplus functions.

For a different environment, however, chapter 4 has presented an opposite result. Given that European competition policy prevents price discrimination, chapter 4 showed that European welfare increases when patent breadth is uniform. The force driving the result was the unequal evolution of consumer surplus and profits in patent breadth. Different patent breadths in order to provide sufficient incentive to innovate are too costly in terms of welfare. By harmonisation of patent protection, the European Community can easily keep profits to the innovator the same while reducing the overall Community static welfare loss.

Another application of the model presented in this chapter concerns the discussion on intellectual property rights in the Uruguay Round, as briefly outlined in section 7.1. The two countries in our model can be interpreted as the northern and the southern region in the world. The north generates many innovations whereas the south generates only few. In terms of our model, one could say that the north has a high and the south has a low innovation density. The model predicts that asymmetric innovation densities lead to extensive patent protection in the innovation-intensive region and narrow protection in the weaker region. This situation can indeed be observed in the world (the north provides strong, the south weak protection) and has caused the contentious

discussion in the GATT Round between north and south. The proposal of northern countries, however, to extend their standards of protection to the south does not Pareto-improve the global welfare if innovation densities are too different (as they seem to be). Moreover, if the strong industry lobby has the effect of weighting profits more heavily than consumers' surplus, the standard of protection will be too high, relative to the global optimum. In addition to the analysis presented here, the next chapter will concentrate completely on this north-south issue of intellectual property protection.

At a special symposium in September 1986 in Paris and Pisa, Uruguay, the GATT countries agreed to put the issue of Trade-Related Intellectual Property Rights on the agenda for the Uruguay Round of Multilateral Trade Negotiations. Previously, the US, supported by the EC and Japan, had made great effort to achieve this. Two earlier attempts of the US to improve international protection, as regulated by the Paris Industrial Property Convention (1883), failed. Both attempts – the formulation of the UNCTAD's Code of Conduct for the Transfer of Technology (1985) and the Adoption Convention of the Paris Convention (1987) inspired by UNCTAD criticism on the anti-economic patent system – eventually missed the required unanimity because the demands of the Group of 77 (developing countries) were not accepted (Cassidy 1989, Koppelman 1989 and Gurruch 1989). These attempts kept circulating in the World Intellectual Property Organization, which replaced the administrative staff of the Paris Convention. This time more support for the proposal to extend global patent protection can be expected in the GATT, where northern countries have more influence, much power, pressure was put to get the issue on the Uruguay agenda.¹¹

The economic argument underlying the pressure of the US, the EC and Japan, and the assistance of the developing countries are presented in Sect. 10. In most discussions two parties are distinguished: the northern and the southern countries. The north advocates extension of its strong standards of intellectual property protection to the rest of the world. The main argument is that if southern exports are not protected against imitation in the north, northern firms forgo profits and, consequently, the incentive to invest in R&D weakens. According to the north, the absence of protection would also harm the southern

1. Besides its accordance with the rules of the Paris Convention, especially it has required it as amended in the form of an agreement, as could be the result of the Uruguay Round, which would be the decision of a special agreement 1994 between the countries of a defined number of states (Koppelman 1989).

Strengthening Worldwide Intellectual Property Protection

At a special ministerial session in September 1986 in Punta del Este, Uruguay, the GATT countries agreed to put the issue of Trade Related Intellectual Property Rights on the agenda for the Uruguay Round of Multilateral Trade Negotiations. Particularly the US, supported by the EC and Japan, have made great effort to achieve this. Two earlier attempts of the US to improve international protection, as regulated by the Paris Industrial Property Convention (1883), failed. Both attempts - the formulation of the UNCTAD's Code of Conduct for the Transfer of Technology (1980) and the Revision Conference of the Paris Convention (1982), inspired by UNCTAD criticism on the international patent system - eventually missed the required unanimity because the demands of the Group of 77 (developing countries) were not accepted (Cornish 1989, Kunz-Hallstein 1989 and Ullrich 1989). These attempts were undertaken via the World Intellectual Property Organization, which performs the administrative tasks of the Paris Convention. Since more support for the proposal to extend global patent protection can be expected in the GATT, where northern countries have more influence, much political pressure was used to get the issue on the Uruguay agenda.¹

The economic arguments underlying the pressure of the US, the EC and Japan, and the resistance of the developing countries are reasonably clear. In most discussions two parties are distinguished: the northern and the southern countries. The north advocates extension of its strong standards of intellectual property protection to the rest of the world. The main argument is that if northern exports are not protected against imitation in the south, northern firms forego profits and, consequently, the incentive to invest in R&D weakens. According to the north, the absence of protection would also harm the southern

1. Besides, in accordance with the rules of the Paris Convention, unanimity is not required if an amendment in the form of an agreement, as could be the result of the Uruguay Round, is to be accepted; the adoption of a special agreement only requires the consensus of a limited number of states (Kunz-Hallstein 1989).

countries because welfare is reduced if northern firms do not export and face weaker innovation incentives. The south is often mentioned as the other party but in fact contains (at least) two groups: the developing countries and the Newly Industrialized Countries (NICs). The aims of these southern subgroups are identical, but for different reasons. The developing countries make the point that strong intellectual property protection only benefits northern firms - if these firms export to them at all - because larger profits can be expected with stronger protection. These profits flow back to the northern home countries, and the only result is that more southern consumer surplus is lost. Furthermore, most innovations are developed primarily for the northern market and are not necessarily appropriate for developing countries, which have their specific wants and needs. In addition, the NICs are not eager to institute effective intellectual property rights because they have comparative advantages in their low labour costs and good imitation capabilities. Profit opportunities for NIC firms only exist if the original northern innovations are only weakly protected.

Several studies on the international economics of intellectual property, both theoretical and empirical, have appeared recently. Feinberg and Rousslang (1990) estimate the economic effects of foreign patent infringement on US firms. Their analysis is based on a study of the US International Trade Commission on foreign infringement of US intellectual property. They show that, due to foreign intellectual property infringements, US firms have foregone profits of about 1% of their total 1986 sales. Most theoretical models on the topic use a technology-gap theory where trade between north and south occurs because of comparative advantage of the north in R&D and technology-intensive products.² Chin and Grossman (1990) study the argument of improved R&D incentives for northern firms if the south strengthens its patent protection. Their model is a duopoly with a northern firm, which alone can reduce costs through R&D, and a southern firm, which can imitate perfectly. If the south protects the process innovation, the northern firm has a cost advantage in the global competition with the southern firm. Both firms compete on equal terms if the south does not provide protection for the process innovation. The interests of northern and southern governments generally conflict. The north always wants extension of protection whereas the south does not, unless it is a significant demander on the world market and, consequently, gains sufficiently from the process innovation. The optimal global protection can go either way. For instance, much global protection is optimal if R&D productivity is high. Deardorff (1991) constructs a model which points at the decreasingly positive effects on innovation incentive as patent protection is

2. See Krugman (1979a) for a typical technology-gap model and Soete (1981) and Soete and Cimoli (1992) for empirical evidence on technology-driven trade.

extended to a larger portion of the world. In a two-country model, the innovating country benefits from larger patent protection but the technology-importing country often loses more. A richer model, applicable to both product and process innovations and allowing for more freedom of choice in the level of protection, is developed by Diwan and Rodrik (1991). Their main argument concerns the appropriateness of technical development for the south. If the south provides protection, it not only affects the intensity but also the direction of R&D, that is toward technologies which are better suited for the south. This may partially offset the tendency for free-riding behaviour of the south. Helpman (1993), finally, studies international intellectual property rights in a dynamic general equilibrium model. One of his conclusions is that the south seldom benefits from tighter protection while for the north it mainly depends on terms-of-trade effects.

This chapter presents a model which incorporates two aspects of optimal global patent protection not dealt with before: (i) *transport costs* occurring in international trade; and (ii) *imperfect imitation* capability in the south. First consider the issue of transport costs. There is a rich tradition in empirical international economics, including most prominently Linneman (1966) and Leamer and Stern (1970), which studies the role of distance in trade. The results generally show that trade flows between two countries decrease as these countries are further apart. The distance between two countries should be interpreted broadly; it not only includes physical distance, measurable in miles, but can also include metaphorical distance, such as differences in culture. Consequently, Armington (1969) makes the observation that distance brings about that each country produces a unique good and that goods in international trade are thus differentiated by country of origin. Embedded in this empirical distance tradition, Ferrantino (1993) presents an empirical analysis of intellectual property rights. In brief, he finds that for US exporters in 1982, distance from trading partners plays an important role. Furthermore, he finds evidence, be it weak, that exports to countries with weak intellectual property protection are lower (the extent of protection is measured by membership of the Paris or Berne Convention on intellectual property and patent durations). In this chapter, a theoretical model is advanced where distance, either physically or metaphorically interpreted, is included and plays an important role.

The second difference between the model presented here and the previous ones is the assumption on imitation in the south. All previous models assume that the southern capability of imitating northern goods is perfect. A study of David (1991) sheds a different light on the topic. His main argument is that the assumption that the south can imitate easily and perfectly is often not justified. Southern imitation capability might be poor indeed. It can therefore be in the

interest of the south to provide intellectual property protection and induce northern firms to export their new products to them. This argument rests on the historical origin of patent systems. The very first patents in early Venice and England were granted to foreign masters in order to make them teach domestic apprentices their arts and skills (see chapter 2). Without these exclusive rights, such skilful masters would never come and teach. Building further on David's argument, this essay studies scenarios in which southern imitation is imperfect.

To be more concrete, this chapter develops a simple locational model of intellectual property rights in which the world is represented as a line from north to south and each country is a point on the line. Like the previous models presented, the image is borrowed from the product differentiation literature in Industrial Organization, starting with Hotelling (1929).³ The 'north' is defined as countries that provide protection and the 'south' as countries that do not provide protection. Innovations only occur in the north, but they can be imitated, though imperfectly, in the south. The value of an innovation diminishes with distance from the innovator, either because international transport costs are incurred in exporting the product, or because the technology becomes less suitable for more distant countries. The analysis aims at determining the optimal point dividing north and south. The model is appropriate to study various intellectual property rights such as trademarks, patents and copyright, but only insofar as these rights can be exploited abroad by exporting products. The model does not cover technology that can be exploited abroad only through licensing or foreign direct investment because such technology is not likely to be imitated in the south. The focus of exposition in this chapter is on patents, rather than on other forms of intellectual property.

The basic model, presented in section 8.1, examines the static welfare effects of strengthening worldwide patent protection and reveals, from this static point of view, the interests of the northern and southern regions and of the complete world. Endogenous innovation enters the analysis in section 8.2. The effects of patent protection on innovation incentives of northern firms are included. Section 8.3 presents some extensions of the basic model. Finally, section 8.4 concludes and points at directions for future research.

3. The first application of location models in a theoretical international trade context is Lösch (1954); later on, Lancaster (1980) was the first using Hotelling-type, or address models, in international trade whereas Krugman (1979b) started the Chamberlinian, non-address paradigm of product differentiation in this context.

8.1 Global Patent Protection with Exogenous Innovation

Consider a linear world, similar to Hotelling's (1929) linear city, which extends from 0 to 1.⁴ Northern countries are located at the 'left-hand side' in this world with the most northern country at 0. Moving to the 'right' means moving southward where the southernmost country is located at 1. Countries are uniformly distributed along this world with density 1 and their location is given by $w \in [0, 1]$. Suppose there is a firm in the northernmost country $w = 0$ (in section 8.3, other resident countries will be considered) which is a monopolist in the invention market of a worldwide industry and has generated a product innovation. This firm wants to export its new product to the rest of the world and in order to protect its exports it seeks patent protection in as many countries as possible. Not every country, however, provides effective patent protection. Roughly representing the situation in the real world, suppose that the countries which do provide patent protection are located in the northern region $[0, b]$ and the countries which do not are located in the south $[b, 1]$, where $b \in [0, 1]$ is the border country without an effective patent system.⁵ In order to simplify notations and verbal descriptions, let the north and the south be synonymous with the regions where effective patent protection is provided and not provided, respectively. The north and the south are treated here as two homogeneous coalitions in which individual countries are not concerned with their national interest but only with the interest of their coalition. I will address this coalition assumption later. The length and scope of protection (breadth and height) within the north are furthermore assumed to be exogenous and uniform.⁶

The consumer surplus generated by the new product in each individual

4. One way of imagining the transformation of the world as a globe into one as a line is the following: collect all countries which are located at the same degree of latitude at one degree of longitude. A location on the line then represents all countries at the same degree of latitude. In the line presentation, distances west- and eastward, between countries at the same degree of latitude, are lost.

5. I am aware that this picture violates the fact that some southern countries such as New Zealand, Australia and South Africa do provide effective protection.

6. An interesting problem, which will not be studied here, would be how the scope of protection that the north wants for the whole world affects the southern decision on patent protection. If the scope of protection is narrow and a patentholder consequently has weak monopoly power (Gilbert and Shapiro 1990 and Klemperer 1990), the south may be more inclined to provide effective protection and license the innovation at better terms. In the case of broad protection, on the contrary, the south may be more inclined to illegal borrowing.

country on the line is given by:

$$U = \begin{cases} v - p - td & \text{if the country imports} \\ 0 & \text{otherwise} \end{cases} \quad (8.1)$$

where v is the gross surplus which the innovation provides (the quality), p is the 'free-on-board' price of the innovation, t is the cost per unit distance of transporting the product over the world ($t > 0$), and d is the distance from the exporting country to the importing country w (in this case $d \in [0, 1]$). A country consumes one unit of the new product if its consumer surplus is non-negative, and no unit otherwise.⁷ Distance in this model can be interpreted literally in miles or kilometres; td is then the total cost of carrying the new product by cargo ship, truck, train or plane, from the innovating firm in country $w = 0$ to the importing country $w = d$. As previously mentioned, there is also an alternative, metaphorical interpretation of the transport costs, which is quite common in the product differentiation literature. A location on the line can be thought of as representing a country's most preferred variety of the new product. Each country has its own preferred variety of the new product, for example because of country-specific parameters such as infrastructure, national consumption pattern, education level and natural environment. The transport cost is *then analogous* to a surplus penalty caused by importing a less preferred variety. According to this interpretation, the new product is most appropriate for the country from where it originates, maybe because the inventor is more familiar with the local circumstances and demand characteristics (compare Diwan and Rodrik 1991 for the point of appropriate technologies). Countries on the line are ordered according to the distance of their preferred variety from the produced innovation. Although this ordering principle differs from the geographical principle, there certainly is a plausible correspondence. Fellow northern countries face relatively small surplus losses, and moving southward these national losses grow larger as the product becomes less and less appropriate. This reflects the fact that the difference between northern and southern countries is larger than the differences within the north or the south.

The countries which do not provide effective patent protection are assumed to imitate the new product costlessly. Despite the absence of patents, which can hinder imitation, southern countries are not able to imitate perfectly; their imita-

7. The assumption that the demand per country is unitary and inelastic of course foregoes the notion of distance within a country and a national demand function.

tions generate less gross surplus, namely pv , with $0 < \rho < 1$.⁸ The whole region $[b, 1]$ is filled up with these imperfect imitations pv (see figure 8.1). By assumption, the innovator cannot make any profits in the unprotected region $[b, 1]$. Imitation prevents this. What are the innovator's profits in the patent-protected region of the world $[0, b)$? Working backward, I will first determine the demand function for the patentholder which is based on the national consumer surplus function (8.1), and then derive the optimal price and (gross) profits. Together with an innovation cost function, these optimal gross profits will be used in the next section to determine the optimal innovation level. This section studies the optimal patent protection for exogenously given innovations. The implications of international patent policy for the static global welfare can be isolated this way.

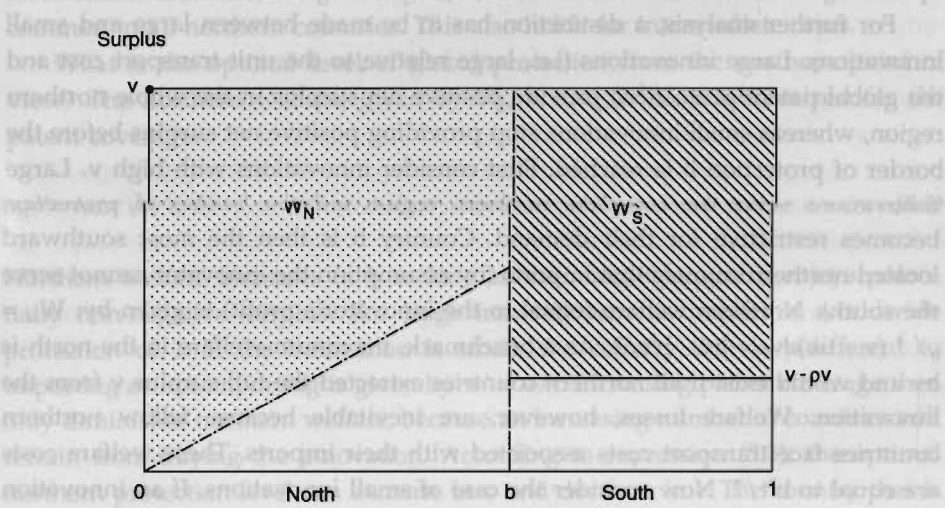


Figure 8.1 Northern and southern welfare

8. The imitation parameter ρ may be a function of the national knowledge base which can be dependent on the average level of education or the regional R&D expenditures. See Verspagen (1991) on the determinants of international spill-overs and national learning capabilities. Taylor (1993), on the other hand, develops a model where the ease of imitation is determined by the original innovator, who can masque his innovation as to make imitation more difficult. Finally, the lower surplus of Southern imitations can be the result of a lower *perceived* quality. All three interpretations can be behind the imperfect imitation capability here.

The innovator is assumed to be perfectly able to price discriminate over countries. This assumption is not decisive for the eventual conclusions; the results are very similar in the case of uniform pricing. The marginal cost of production is set equal to 0 for convenience. The patentholder maximizes his profits by choosing a price which totally absorbs the consumer surplus in the northern region:

$$p^* = v - tw \quad (8.2)$$

The profit of the patentholder is depicted in figure 8.1 by the dotted area. Each country $w \in [0, b]$ is prepared to pay the value of the innovation v minus the transport costs tw that must be incurred to import the innovation. Since the northern consumer surplus flows entirely to the patentholder, his profits make up the total social welfare in the north (W_N).

For further analysis, a distinction has to be made between large and small innovations. Large innovations (i.e., large relative to the unit transport cost and the global patent protection) provide positive net surplus in the whole northern region, whereas small innovations stop providing positive net surplus before the border of protection b is reached. First consider innovations with high v . Large innovations serve the complete northern region and the border of protection becomes restrictive for their demand. Country b is then the most southward located northern country that imports (by assumption the innovator cannot serve the south). Northern welfare (equal to the innovator's profit) is given by: $W_N = \int_0^b (v - tw)dw = bv - b^2t/2$. As a benchmark, maximum welfare in the north is bv and would exist if all northern countries extracted the full surplus v from the innovation. Welfare losses, however, are inevitable because fellow northern countries face transport costs associated with their imports. These welfare costs are equal to $b^2t/2$. Now consider the case of small innovations. If an innovation v is small, the country that is indifferent to importing or not importing is given by $w' = v/t$. The countries north of w' will thus not consume the product. Northern welfare is then given by $W_N = \int_0^{w'} (v - tw)dw = v^2/2t$.

The welfare in the southern region (W_S) without patent protection is totally made up of consumer surplus as depicted by the shaded area in figure 8.1. Each southern country can imitate and will thus incur no transport cost. Yet welfare losses will occur because surplus is lost while imitating the original innovation. For both large and small innovations, southern welfare is $W_S = (1 - b)pv$.

Based on the welfare functions for north and south, it is possible to determine the optimal level of global patent protection. Before examining the optimal global level, however, I will first study the interests of the northern and southern regions separately. As mentioned above, north and south are treated as coalitions.

tions, where individual countries do not pursue national interests. In the simple scenario studied here, with only one innovator resident in the most northern country, one could object that fellow northern countries ignore their national interests as their consumers' surpluses flow away to the patentholder and for some of them imitation would yield increased national welfare. Two arguments can be advanced as a defence. First, the assumption of homogeneous coalitions is inspired by the practice of the GATT negotiations, where indeed two blocks can be found. The northern block represents the interests of innovators in general because, unlike the simple scenario studied here, innovators are likely to be resident in all these countries. Second, even in this simple scenario, there is a possible justification for the existence of a homogeneous northern coalition. Suppose that the patentholder compensates all countries that provide protection. In the extreme case that the innovator pays an amount equal to his total profit to all northern countries, the goal of profit maximization remains unchanged and is common to all northern countries. This also holds for intermediate cases.

What is the optimal level of global protection from the northern point of view? The northern welfare, given by $W_N = bv - tb^2/2$, is maximized for the patent coverage:

$$b_N^* = \min(v/t, 1) \quad (8.3)$$

Northern welfare increases in b as long as the marginal country enjoys potentially non-negative surplus. For large innovations, this implies a worldwide protection of 1. If the innovation is small, the country that is indifferent to importing or not importing is given by $w' = v/t$. Increasing patent coverage then only diminishes northern welfare, because an increasing number of countries will refrain from buying the innovation. According to expression (8.3), the optimal northern protection level can increase in v and decrease in t . The country that is indifferent to importing or not importing is located more southward if the innovation provides more surplus. The alternative of not importing is also more attractive if the transport costs are larger.

The welfare in the southern region without patent protection is totally made up of consumer surplus. Southern welfare always decreases in world patent coverage ($dW_S/db < 0$ for $b \in [0, 1]$) because the region shrinks without getting better innovations in return. The negative effect of extending global protection intensifies as ρ and v become larger. So the optimum from the southern perspective is to provide no protection at all (assuming that the innovator still generates his innovation so that imitation continues to be possible).

How do these optimal levels of the north and the south, which can be extreme opposites (none vs. total protection), combine in one optimal global level

of protection? The unweighed sum of the northern and southern welfare is taken as the global objective function. The optimal extent of protection that maximizes the objective function is:

$$b_G^* = \min (v(1 - \rho)/t, 1) \quad (8.4)$$

Notice that this optimal global protection is smaller than or equal to the optimal one from the northern point of view (given by 8.3), and always larger than the optimal southern level. An interesting property of the optimal global level is that it decreases in the imitation capability of the south. The reason for this is that the south loses more when it provides patent protection if its imitation capability is high.

8.2 Endogenous Innovation

So far, the analysis has revealed the interests of the north and south from a static perspective, i.e., for given innovations. This section will take into account the effects of protection on the R&D incentive of the innovator. The innovation v is chosen endogenously by the innovator. His choice of R&D expenditures is a function of the global patent protection provided. On the basis of the gross profits (net of R&D costs) which can be gained with a certain innovation, the inventor has to choose an innovation level. Suppose that by increasing R&D expenditures a higher gross surplus is being generated by the product innovation, and suppose that this is decreasingly so. A simple deterministic innovation cost function which catches this idea of exhausting innovation opportunities is:

$$c(v) = \alpha v^2 \quad \alpha > 0 \quad (8.5)$$

where c stands for the R&D expenditures and $0 \leq v \leq v^*$, where v^* is the upper limit of v (the final improvement possible). For small innovations - the ones that do not serve all countries - the net profit function is $\pi(v; p^*) = v^2/(2t) - \alpha v^2$. If the innovation costs are relatively high ($\alpha \geq 1/(2t)$), then the innovation will not be developed at all. If the innovation costs are relatively low ($\alpha < 1/(2t)$), which I will focus on from now on, the innovator always gains with investing more R&D. The upper limit v^* will occur when the innovation opportunities are completely exhausted before the border of patent protection is reached ($v^* < tb$). The market will then not be completely served and the global patent protection provided does not affect the innovation incentive. For these innovations, the optimal global protection is identical to the one that maximizes static welfare $b_G^* = \min (v^*(1 - \rho)/t, 1)$, as given by (8.4). However, if the upper limit on surplus is

relatively high ($v^* \geq tb$), we enter the regime where the most southward importing country is located at the border b . For this regime, the net profit function (gross profit minus R&D cost) is $\pi(v; p^*) = bv - tb^2 - \alpha v^2$. The optimal innovation level which maximizes this net profit function is:

$$v^* = b/(2\alpha) \quad (8.6)$$

This expression indicates that a higher unit innovation cost α makes the optimal level of innovation lower. Furthermore, the optimal innovation is positively related to the level of protection. So the claim of the northern countries in the Uruguay Round that strengthening world protection improves the R&D incentive of northern firms is indeed captured by the model. Again, the welfare in the north is given by the profits of the discriminating monopolist: $W_N = bv^* - tb^2/2 - \alpha v^{*2}$. The extent of patent coverage b has opposing effects on northern welfare. The positive welfare effects are, first, that increasing coverage improves the induced innovation (v^* increases in b) and, second, but this is due to the terminology used, that the size of the northern region increases. There are also negative effects on northern welfare. First, similar to the static analysis, an increase of b causes larger transport costs. Second, since the induced innovation is larger, the R&D costs are relatively higher. The gain in welfare always turns out to be large enough to compensate for the extra transport and R&D costs. To see this, substitute the optimal innovation level (8.5) in the expression for welfare which is then reduced to: $W_N = b^2(1 - 2\alpha t)/(4\alpha)$. Since the current analysis is based on relatively low innovation cost ($\alpha < 1/(2t)$), northern welfare always increases (quadratically) in b . Therefore, the optimal protection from the northern point of view is the corner solution $b_N^*(v^*) = 1$.

Again the welfare in the southern region $[b, 1]$ is totally made up of consumer surplus. For the optimal development level v^* , the southern welfare is:

$$W_S = (1 - b)pv^* = (1 - b)(pb/2\alpha) \quad (8.7)$$

Contrary to the previous section, the south faces a trade-off now. More patent protection enhances the R&D incentive, which improves the product innovation and, consequently, the southern imitations on the one hand, but causes welfare loss due to the decreased size of the southern region on the other. From the southern point of view the optimal patent coverage is:

$$b_S^*(v^*) = 1/2 \quad (8.8)$$

'Equatorial' protection is now in the best interest of the south. Compared to the

case of exogenous innovation discussed in the previous section, it is now in the interest of the south to provide some protection in order to induce sufficient innovation incentive.

The global welfare is again the unweighed sum of northern and southern welfare, $W_G = bv^* - tb^2/2 + (1 - b)pv^* - \alpha v^{*2}$, and is maximized for:

$$b_G^*(v^*) = \begin{cases} 1 & \text{if } \rho \leq 1 - 2\alpha t \\ \rho/(2\rho + 2\alpha t - 1) & \text{if } \rho > 1 - 2\alpha t \end{cases} \quad (8.9)$$

Solution (8.9) has some important properties for international patent policy. If the imitation capability in the south is poor ($\rho \leq 1 - 2\alpha t$), a benevolent global social planner provides worldwide patent protection of 1. Apart from imitation capability ρ , the size of the regime of complete coverage also depends on the unit innovation transport cost t and the innovation cost α . This regime enlarges if the unit transport cost decreases, relative to the innovation cost ($1 - 2\alpha t$ is larger then). Thus, innovations that cause relatively low transport costs, i.e., whose value decreases slowly when moving southward (for example, technologies that are appropriate for all countries), should be given worldwide protection. A regime of non-worldwide protection emerges if the south is better able to imitate ($\rho > 1 - 2\alpha t$). From the optimal protection level onwards, an extra southern country that provides protection adds insufficiently to the innovation incentive in order to make up for the difference in net surplus between the original innovation and the imitation: in this regime $2\alpha t - 1 < 0$, the optimal global protection is always larger than $1/2$. Table 8.1 presents some numerical examples. The table shows that the globally optimal levels of patent coverage are higher if ρ is lower. It furthermore shows that, irrespective of ρ , the optimal levels are lower if the transport cost, relative to the innovation cost, is higher. This can be the case for a high absolute transport cost but also for a low absolute innovation cost.

Table 8.1
Optimal global patent coverage

Transport cost t	b_G^* ; $\rho = 0.8$	b_G^* ; $\rho = 0.6$	b_G^* ; $\rho = 0.4$	$b_G^* < 1$
$1/(10\alpha)$	1	1	1	$\rho > 0.82$
$1/(4\alpha)$	0.727	0.857	1	$\rho > 0.50$
$1/(3\alpha)$	0.632	0.692	0.857	$\rho > 0.33$
$1/(2.1\alpha)$	0.515	0.521	0.532	$\rho > 0.06$

The simple model developed in this section has thus demonstrated that the aim of northern countries in the Uruguay Round to extend patent protection globally does not necessarily need to be welfare improving from the global point of view. There are two reasons why the optimal level of protection is not worldwide protection. First, if southern capability to imitate northern innovations is high, the welfare emerging from southern imitations is higher than the welfare emerging from stronger innovation incentives. Second, if the unit transport cost is relatively high (in other words, if the innovation is only appropriate for the innovator's direct neighbourhood), worldwide protection would provide too strong an incentive to the innovator and overcompensate him. Although these conclusions are intuitively clear, one should keep in mind that they are based on a simple model which focuses only on imperfect imitation and transport costs. Of course, more serious policy conclusions should be based on more realistic models. But the result that worldwide protection does not need to be optimal, as northern countries generally claim, should not be ignored in the Uruguay Round negotiations.

8.3 Some Extensions of the Basic Model

To relax some of the simplifying assumptions this section will study four extensions of the basic model. The first extension (a) examines locations of the innovator other than those in the basic model where the innovator was resident in the northernmost country. The second extension (b) studies the effect of an uneven international income distribution. The third extension (c) determines the impact of the assumption of the south being completely filled with imitations and looks at what happens if there is only one imitator (for example, a NIC). The final extension (d) analyzes the 'exhaustion doctrine', which is often practised in international jurisdiction.

(a) Innovator Resident in Another northern Country

Up to this point, the innovator has been a resident of the northernmost country $w = 0$. Suppose now that the innovator is a resident of a country $w^* \in [0, b]$. What is the effect of the innovator's location? Suppose that the complete northern region is served ($\alpha < 1/(2t)$, $v^* > \max(t(b - w^*), tw^*)$). The profit of the patentholder is then given by $\pi = \int_0^{w^*} (v - t(w^* - w))dw + \int_{w^*}^b (v - t(w - w^*))dw = bv - t(b^2 - 2bw^* + 2w^{*2})/2$. The optimal innovation based on this gross profit function and the R&D cost function (8.4) is independent of the innovator's location: $v^* = b/(2\alpha)$. Relative to the basic analysis in section 8.2, the profit of the innovator and, therefore, the northern welfare has increased. The reason is that a

more central location of the innovator causes fewer transport costs.⁹ The optimal levels of protection from both regional perspectives remain unchanged: $b_N^* = 1$ for the north and $b_S^* = 1/2$ for the south. The increased welfare in the north, however, does show up in the optimal level of protection from the global point of view:

$$b_G^* = \begin{cases} 1 & \text{if } \rho \leq 1 - 2\alpha t(1 - w^*) \\ (\rho + 2\alpha w^*t)/(2\rho + 2\alpha t - 1) & \text{if } \rho > 1 - 2\alpha t(1 - w^*) \end{cases} \quad (8.10)$$

The optimal level is larger if the innovator is located more southward. The reason is that the marginal effect on the innovation incentive of an extra southern country providing patent protection is larger if the innovator is located more southward. The qualitative results of the basic model remain unchanged: non-worldwide protection is optimal if the imitation capability is good and the unit transport cost t is large, relative to the innovation cost α .

(b) Differences in National Income

Extension (a) might indicate that firms located in the centre of the northern region are more innovative since they can gain more from an innovation. The notion that there are differences in national income over the world and that southern countries are often poorer might offset this tendency. The national consumer surplus function as given by (8.1) can be extended such that differences in national incomes between countries are incorporated. Consider the following national consumer surplus function:

$$U = \begin{cases} mv - p - td & \text{if the country imports} \\ 0 & \text{otherwise} \end{cases} \quad (8.11)$$

where m is a country-specific parameter which indicates the degree to which the country appreciates the innovation ($m \in [y, z]$, $z \geq y$). This innovation preference intensity parameter m can be interpreted as the marginal rate of substitution between national income and importing the innovation. Consider, for example, (see Tirole 1989, p.97) the separable utility function $U = u(I - p) + v$, where I is national income. Let I be much larger than p . A first-order Taylor expansion of this utility function is: $U = -u'(I)p + v$. Define $m = 1/u'(I)$. If u is concave, wealthier countries have a low $u'(I)$ and therefore a higher m .

9. The best location for the innovator, from the Northern point of view, would be the centre of the Northern region $w^* = b/2$. The total Northern transport costs $t(b^2 - 2bw^* + 2w^{*2})/2$ are minimized then.

It is commonly known that national incomes are not uniformly distributed over the world. northern countries are generally richer and southern countries poorer. In order to give (extreme) credit to this fact, let the northernmost country $w = 0$ be the richest ($m = z$) and moving southward countries become poorer with the poorest country ($m = y$) being the southernmost one at $w = 1$. In other words, let $m = z - (z - y)w$. Figure 8.2 illustrates the income differences for $z = 1$ and $y = 0$ and shows that the potential national surpluses decrease while moving southward. Again, the innovator is located in the northernmost country and chooses country-specific discriminating prices.¹⁰

The profit-maximizing price charged to country w is:

$$p^*(w) = mv - p - tw = (z - (z - y)w)v - v \tag{8.12}$$

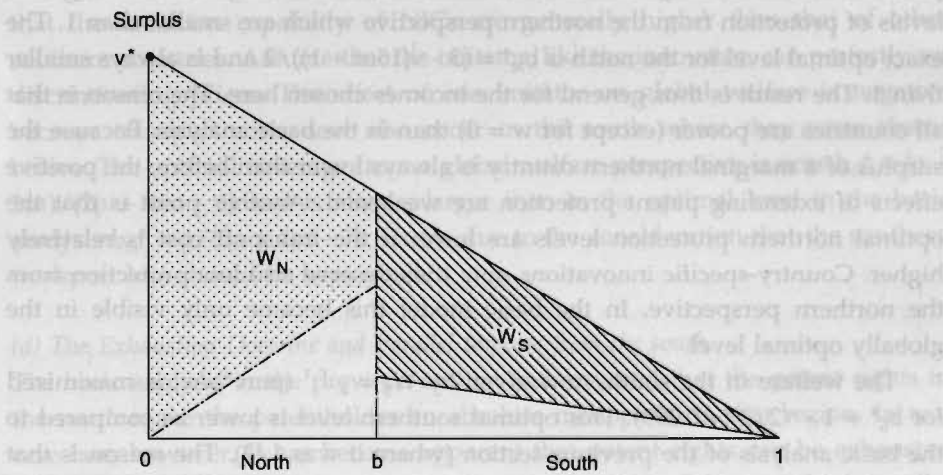


Figure 8.2 Differences in national income

10. Notice that profitable arbitrage can occur here. Consider two countries with different income, m_1 and m_2 , with $m_1 > m_2$. Country m_2 can resell the new product to the richer country m_1 (which appreciates the innovation more than country m_2 itself) and make profits of $v(m_1 - m_2)$. The practice of arbitrage between countries, however, is excluded from the analysis here.

Let the northern region be served completely. The patent coverage is then restrictive and the profits of the patentholder are given by: $\pi = \int_0^b ((z - (z - y)w)v - tw)dw = zvb - (vb^2(z - y))/2 - tb^2/2$. Based on these gross profits and the R&D cost function, the optimal innovation level is:

$$v^* = (2zb - (z - y)b^2)/(4\alpha) \quad (8.13)$$

Expression (8.13) has some interesting properties. The optimal innovation level v^* is larger if the richest country z is richer and smaller if the difference between the richest and the poorest $(z - y)$ is larger. In fact, $(z - y)$ represents the skewness of global income distribution (since the world line stays of length 1). The levelling of incomes thus has a positive effect on the innovation v being chosen.

In order to prevent that the expressions for the optimal levels of protection become untransparently complex, I set the income parameter y equal to 0 and z equal to 1. Table 8.2 presents the optimal coverage for various constellations of the unit transport cost t and the southern imitation capability p . Contrary to the previous results for large innovations, table 8.2 (second column) shows optimal levels of protection from the northern perspective which are smaller than 1. The exact optimal level for the north is $b_N^* = (3 - \sqrt{16\alpha t + 1})/2$ and is always smaller than 1. The result is thus general for the incomes chosen here. The reason is that all countries are poorer (except for $w = 0$) than in the basic analysis. Because the surplus of a marginal northern country is always lower than before, the positive effects of extending patent protection are weakened. Another point is that the optimal northern protection levels are lower if the transport cost is relatively higher. Country-specific innovations now thus deserve also less protection from the northern perspective. In the basic model this became only visible in the globally optimal level.

The welfare in the south, now given by $W_s = \int_0^1 (pmv^*)dw$, is maximized for $b_s^* = 1 - \sqrt{2}/2 (\approx 0.293)$. This optimal southern level is lower as compared to the basic analysis of the previous section (where it was $1/2$). The reason is that northward expansion is welfare-improving for a longer period of time as richer countries join the region and extract more surplus from imitation. Table 8.2 reveals some interesting aspects of the optimal level of protection from a global perspective. As before, the optimal global levels are lower if the unit transport cost is higher, relative to the innovation cost. However, as opposed to previous results we also see that the optimal level of protection can decrease if the southern imitation capability is worse. This is the case for high transport costs ($t = 1/(3\alpha)$ and $t = 1/(2.1\alpha)$). Relatively high transport costs make the wasteful transport loss large to the extent that, even inferior, imitations are the best alternative.

Table 8.2Optimal coverages with national income differences ($z = 1, y = 0$)

Transport cost t	b_N^*	b_S^*	$b_G^*; \rho = 0.8$	$b_G^*; \rho = 0.6$	$b_G^*; \rho = 0.4$
$1/(10\alpha)$	0.694	0.293	0.477	0.527	0.586
$1/(4\alpha)$	0.382	0.293	0.322	0.329	0.339
$1/(3\alpha)$	0.242	0.293	0.278	0.275	0.270
$1/(2.1\alpha)$	0.032	0.293	0.227	0.215	0.198

(c) One Imitator in the south

The poorest developing countries can hardly be expected to imitate, even imperfectly. Instead of imitators in all southern countries, suppose now that there is only one imitator in the south, resident in country b . This country could be thought of as being a NIC. Extension (c) then reflects (be it extremely) the fact that the imitation capability of NICs are generally higher than that of other southern countries. Assume that the imitator, like the innovator, can perfectly set discriminating prices. The effect of one imitator on global welfare is negative because transport costs now also occur in the south where they were absent before. The optimal coverage from a global welfare perspective is now $b_G^* = (\rho + 2\alpha t)/(2\rho + 4\alpha t - 1)$ and is always larger than in the optimal level in the basic analysis $b_G^* = \rho/(2\rho + 2\alpha t - 1)$. This is due to the consideration that the southern transport loss should be kept small.

(d) The Exhaustion Doctrine and Parallel Imports from the south

The basic analysis in the previous section has assumed that the patent rights in the north make the patentholder an absolute monopolist in that region. In real case law, however, it has become apparent that patent rights can be exhausted under certain circumstances. For instance, in the case of lower-priced parallel imports. Consider, in figure 8.3, the country which is located just a little north of the border of protection b and faces a considerable (delivery) price differential between the original northern product and the southern imitation, as illustrated by Δp . A court in this country may easily interpret the price differential as being a signal of abuse of monopoly power and hold the patentholder's rights to be exhausted. The main argument for this *exhaustion doctrine* is that it promotes the free trade of new products and sets (indirect) limits to the monopoly price of a patentholder. The patentholder always has to take into account, when choosing his price, that foreign competitors can be permitted to enter. Judging from inter-

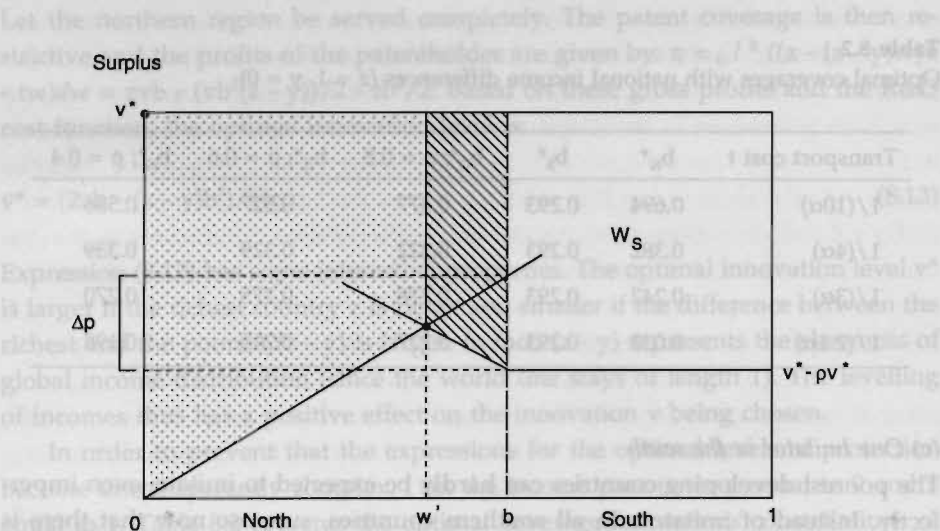


Figure 8.3 The exhaustion doctrine

national case law, the exhaustion doctrine is widespread. In the US and the EC, for example, it is common practice (Cornish 1989).

In terms of the basic model, the exhaustion doctrine implies that patents only provide protection against imitations or duplications which are made within the patent-granting country, but not against cheaper imitations made abroad. The patentholder actually faces import competition from the south now and must compete in the region $[0, b)$ with the closest imitation from country b. In the northern region two firms are now competing: the innovator and the imitator from b. The imitator has an advantage in the southern part of the north because countries located there face higher transport costs when buying from the original innovator. The advantage of the original innovator is that his product provides more surplus. Again assume that the imitation industry in the south is competitive (imitators are present in each country) and that the imitator in b charges the competitive price of $p = 0$. The country which is indifferent to buying from the original innovator at the (perfectly discriminating) price $p^* = v - tw$, or from the imitator in b is given by $w' = (v(1 - p) + tb)/(2t)$ (see figure 8.3). No country to the south of w' buys the original innovation because the surplus from the imitation is larger. Countries to the north of w' all buy the original innovation at price

$p^* = v - tw$.¹¹ The gross profits of the patentholder are given by: $\pi = \int_0^{w'} (v - tw)dw = (v(1 - \rho) + bt)(3v + \rho v - bt)/(8t)$. Based on these gross profits and the innovation production function, the optimal innovation level is: $v^* = bt(1 + \rho)/(\rho^2 + 2\rho + 8\alpha t - 3)$. The welfare in the north now includes the profits of the patentholder, illustrated by the dotted area in figure 8.3, and the consumer surplus in the countries south of w' that buy an imitation at a competitive price from country b, illustrated by the shaded area. The welfare in the southern region does not change compared to the basic scenario.

Table 8.3
Optimal patent coverage with and without the exhaustion doctrine ($\rho = 0.8$)

Transport cost t	Exhaustion Doctrine	No Exhaustion Doctrine
$1/(10\alpha)$	0.120	1
$1/(8\alpha)$	0.502	0.941
$1/(4\alpha)$	0.806	0.727
$1/(3\alpha)$	0.746	0.632
$1/(2.1\alpha)$	0.626	0.515

Table 8.3 compares the globally optimal coverages under the exhaustion doctrine with the ones from the basic analysis. A surprising result emerges for relatively low transport cost ($t = 1/(10\alpha)$ and $t = 1/(8\alpha)$).¹² Under the exhaustion doctrine, the optimal global patent coverage is smaller than in the basic case. The effects of strengthening protection are somewhat different compared to the basic case. The positive effect of a larger b on the profit and innovation level is

11. Notice that, theoretically, the full discriminating price p^* is not correct since countries North of w' may no longer have a reservation price of 0 (the alternative of not buying). Their reservation price is the net surplus of the imitation from the South ($\rho v - t(b - w)$). The true discriminating price for country w is then: $p = \min (v - tw, v - tw - (\rho v - t(b - w)))$. Although theoretically incorrect, in order to simplify calculations I do not take into account these new reservation prices. The effect is that the profits of the innovator are overestimated and that, consequently, the North gets too much weight in the global welfare function. This might lead to optimal coverages which are too high. Qualitatively, however, the results are very similar.

12. I checked whether for $t = 1/(10\alpha)$ and $b = 0.120$ the net surplus of the indifferent country w' is still positive. This is indeed the case.

weakened because part of the increasing demand now goes to the imitator. The negative effect, however, of a larger b on the transport losses is also weakened. For relatively low transport costs (or high innovation cost), even without much protection, the negative effect is relatively stronger. For higher costs, the results of the basic analysis are confirmed and the optimal levels decrease in t .

8.4 Conclusions

A simple address model of global patent protection is developed which mainly shows that the proposal of the US, the EC and Japan, during the Uruguay Round, to extend protection worldwide, does not need to be beneficial for the global welfare. Strengthening worldwide protection is likely to be beneficial if the imitation capability in the south is poor. The south can then better provide patent protection, thereby improving the northern innovation incentive, and import the original innovation rather than producing a poor imitation itself. If, however, the imitation capability in the south is better, extending protection might worsen global welfare. In the latter case, an imitation provides more surplus than the original innovation which involves transport costs. The level of optimal global protection in general decreases for higher transport costs (relative to innovation costs). Thus, innovations that are only locally appropriate should not be given worldwide protection. Several extensions of the basic model have been examined but in general do not change the main message very much.

It should be stressed that the above conclusions follow from a simple model with simplifying specifications. The transport cost function, the R&D cost function and the distributions of consumers and producers could be modelled more generally. Particularly the incorporation of non-unitary and elastic national demand curves may lead to richer conclusions. Besides these refinements, some other applications of the model also stay open to future research. Licensing and foreign direct investment, for example, make the transport costs decrease and, consequently, the welfare costs of extending global protection will be lower. Another application concerns the trade-off between extending patent protection to more countries and improving the protection within countries. This trade-off is very similar to the one dealt with in Gilbert and Shapiro (1990) and Klemperer (1990) on the optimal mix of patent length and breadth (see chapter 2). In some cases it might be more efficient to provide infinitely long protection in one country and no protection in others, while in other cases shorter lived rights in the whole world might be optimal.

Summary, Conclusions and Directions for Future Research

Although this book does not contain one but several studies, this concluding chapter will try to give a comprehensive overview of the results obtained in the separate essays. The order of presentation as outlined in the Introduction (chapter 1) is maintained: Starting at the industry level, I will work my way up to the country and world level.

First, I want to emphasize again the practical importance of the economics of patent protection. In addition to the numerous cases of patent infringement described in the legal literature (which indicate that protection is imperfect), the empirical observation which supports the relevance of all studies in this book is the fact that most patentable inventions are indeed patented (fact (a) in the Introduction). Competition in industries where invention takes place is thus often affected by the presence of patent protection. Patents create asymmetries in the competition between a patentholder, who enjoys some extra market power, and his competitors, who face restrictions in their strategy spaces. As opposed to the economics of patent protection studied in this book, another branch of Industrial Organization deals with the economics of patent design. This branch examines how patents affect the pre-patent stage of technological competition: What kinds of incentive to carry out research and development do patents provide to private firms? Empirical studies (for example, Taylor and Silberston 1973, Mansfield 1986), however, have demonstrated that the impact of patents as innovation incentives is low for most industries and high for only few industries. At the same time, these studies have shown that in all industries most of the patentable inventions were actually patented. Thus, for most industries the practical relevance of the economics of patent *protection* seems to be higher than that of the economics of patent *design*.

Within the protection literature, the distinction of several dimensions of protection turns out to be very useful. So far, the legal duration of patents has received considerable attention. Other dimensions of patent protection, however,

which say something about the practice of inventing around, seem to be far more important than duration and have therefore been examined in this book.

Summary

The introducing chapter explains that the motivation behind this book is the lack of attention in economics for the evident and close link with patent law and legal practice. This chapter furthermore summarizes some facts about patent protection which have inspired all the analyses carried out in this book. Following the Introduction, chapter 2 presents a survey of the current literature on patents. The survey differs from existing surveys on the topic in the sense that it only reviews the subgroup of theoretical models that have appeared in *Industrial Organization*. Chapter 2 thus not only yields an overview of the field and the state of the art, but also detects some important niches in the patent research spectrum. Based on these findings, the next chapters focus on new dimensions of patent protection.

Chapters 3, 4 and 5 all examine dimensions - one could label them as 'exclusivity' dimensions - that determine the extent of patent protection and the opportunities for inventing around. Chapter 3 studies the *height* of patent protection, a dimension not studied before. The stringency of novelty requirements that patent offices and courts use in their judgement about infringement disputes and patentability of inventions defines the height of protection provided to a patentholder. The effect of patent height is studied in a duopoly where firms compete in product improvements. As the relation between a large and a small improvement is comparable with the relation between high and low quality, a model of vertical product differentiation is used to describe the competition in improvements. The patent owner exercises monopoly power in the portion of the vertical product spectrum that is protected by the patent. The requirement of minimal steps of improvements, determined by the height, limits the strategy space of competitors who want to invent around a patent; competitors are only allowed to locate in the non-protected region of the spectrum. In this vertical differentiation model with asymmetric strategy spaces, it is shown that low patents do not affect the natural market equilibrium without patents. A patentholder can lose with medium patent heights but becomes a pure monopolist if patents provide high protection. The non-patentholder can gain with medium heights but is increasingly worse off with higher patents.

Chapter 4 deals with the dimension of patent *breadth*. Whereas height indicates the degree of protection against *improvements*, patent breadth gives the degree of protection against *imitations*. Imitations can be considered as horizontal differentiations of the patented product (i.e., some consumers prefer the imitation, others the original patented product), while, as mentioned previously,

improvements are vertical differentiations (i.e., at equal prices, all consumers prefer the improvement to the original product). Thus, whereas patent height defines the strength of protection against competition in vertical differentiations, breadth defines the degree of protection against competition through horizontal differentiations. Since competition through imitations is basically different from competition through improvements, the distinction between height and breadth is a useful one.

The focus in chapter 4 is not completely on patent breadth. A patent breadth model is used as a foundation of the demand functions faced by a patentholder. The central object of study is the practice of price discrimination. More precisely, the analysis concentrates on the effect of third-degree price discrimination on dynamic efficiency. The ratio of static welfare loss to profits, which indicates how efficiently an innovation incentive in the form of a minimum profit level can be provided, is taken as a measure of dynamic efficiency. Whereas uniform pricing with linear demand curves improves static efficiency (a well-known result in the discrimination literature), price discrimination is equally efficient to uniform pricing when evaluated from a dynamic point of view. Moreover, if an explicit model of patent breadth underlies the demand curves of the monopolist, it is shown that dynamic efficiency improves with price discrimination. This result is an important exception to the rule of thumb often practised in competition policy (inspired by the theoretical literature), that price discrimination is welfare-inferior to uniform pricing. A direct application of this result is the European competition and patent policy. While national patent breadths may differ substantially, full exploitation in the form of price discrimination is forbidden by competition policy. From the perspective of providing innovation incentives to private firms, the current harmonisation of national patent systems, through the European patent system and the future Community patent system, is welfare improving.

The dimensions of breadth and height are integrated in chapter 5. Starting with the original utility function that Hotelling (1929) used to model horizontal differentiation, I add a preference intensity parameter to catch vertical differentiation. The only other study that combines horizontal and vertical differentiation that I know of is Neven and Thisse (1990), although they use quadratic transport costs instead of the linear transport costs used by Hotelling. The new model in chapter 5 has been developed to allow for an integrated analysis of patent breadth and patent height. As in chapter 3, I examine a duopoly with a patentholder and a competitor and I focus on one product innovation. Building further on a basic innovation, subsequent innovations can take the form of improvements or imitations of this basic innovation (or a combination of both). The basic innovation is patented. For simplicity, further patents are excluded from analysis. Being the owner of the sole patent, the patentholder is completely free to choose

the improvement level that maximizes his profit. The competitor, however, can be restricted in his choice of an improvement. Although there is a natural tendency to create distance between improvements and thus relax price competition, the marginally increasing R&D costs offset this tendency from a certain point. The freely chosen improvement of the competitor may be sufficiently novel not to infringe the patent. If this is the case, the patent does not affect the market equilibrium. If, however, the height of the patent is restrictive, the competitor must choose a larger than optimal improvement which causes his profit to decrease. The other forms of inventing around, imitations, are taken exogenously. By an ordering assumption on the horizontal product spectrum, the patentholder cannot choose another variety of the basic innovation (the spectrum runs from 0 to 1 and, by definition, the patentholder's variety is located at 0, and moving toward 1 imitations are increasingly different from the variety at 0). The competitor cannot choose his variety either; he is assumed to have an imitation at the border of what is marginally allowed (determined by patent breadth), closest to the patentholder. Given the restrictions caused by patent breadth and height, the optimal inventing-around strategy for the competitor can either be an *improvement* strategy ('above' the patent) or an *imitation* strategy ('aside' the patent). The dimensions of breadth and height can actually give direction to these inventing-around strategies. For some parameter constellations, I find that higher patent protection induces imitation strategies and broader protection improvement strategies.

In the theoretical model of chapter 5, the optimal circumventing strategy depends on demand and cost parameters as well as on the patent office standards of breadth and height. In the practice of most industries, two broad categories of innovation strategies can be distinguished which are very similar to the above improvement and imitation strategies. Freeman (1982) distinguishes between *pioneering* and *imitative* innovation strategies. Pioneering strategies, similar to improvement strategies, rely more on basic research. Since basic research explores more novel and unknown technical paths, the inventions emerging from it meet the novelty requirements of the patent office more easily. Indeed, one can find that firms that pursue pioneering strategies often use patents as primary means of protecting the fruits of their basic research (see Freeman 1982). Imitative, or defensive innovation strategies, on the contrary, rely more on applied research and experimental development. Modifications of current (perhaps patented) products in order to create horizontal distance is the main goal of the R&D carried out by defensive firms. These modifications pass the patent office less easily since the requirements of an inventive step and non-obviousness are often not met.

Chapter 6 presents an empirical analysis of these two innovation strategies.

The *success ratio* of a firm, defined as the proportion of all applications filed that pass the patent office and are granted a patent, is the variable to be explained. The basic hypothesis underlying this empirical analysis is that the success ratio of firms with pioneering innovation strategies is higher than that of firms with defensive strategies. Two data sets are used to test this hypothesis. First, a set published by the European Patent Office which contains all applications filed in the period 1978-1993 (August), and the procedural status of these applications. From this set the success ratio of an individual firm can be obtained. The second set contains information about one industry: the data-processing industry. It covers the 100 largest firms worldwide (by 1990 revenues) for the period 1986-1990. Although this set contains very detailed information, direct figures on the amount of R&D spent on basic research or on applied research and development are not available. Therefore, I construct instrumental variables which indicate to what degree firms carry out basic research. Especially two variables are important in explaining the interfirm success ratios. The first is diversification. Since more diversified firms can spread the risks better and apply the results to a broader extent, diversification is an indication of the degree of basic research a firm is expected to conduct. The second variable is the propensity to patent, defined as the patents to R&D expenditures ratio. Since basic research generally leads to more patents (relatively per unit of R&D expenditure), the propensity to patent is positively related to the success ratio. In the estimations carried out, both these variables are in general highly significant in explaining the success ratios. The basic hypothesis is thus largely confirmed.

In addition to the results on success ratios, the study reveals a remarkable practice of the European Patent Office: Japanese and US applicants have to wait longer for their files to be processed than European applicants. Finally, I find that countries which have more stringent national novelty requirements (represented by a lower proportion of applications passing the national patent office) generally perform better in the European Patent Office. More applications originating from these countries pass the office. National patent offices may thus act as screening institutes for the European Patent Office.

After examining patent dimensions at the industry level in the first chapters, the final two chapters deal with the international patent policy setting. Chapter 7 presents a two-country model of patent breadth. Drawing on Hotelling's model of spatial competition, an address model is developed which makes it possible to study normative and positive aspects of the choice of patent breadth in an international setting (i.e., for two countries). The following results are established. First, when both countries place equal weight on profit and consumers' surplus of their own citizens, relative to the social optimum, patent breadths are too narrow. This result reflects the existence of a positive externality flowing from

each country's patent breadth to the profit and consumers' surplus enjoyed by citizens of the other country. Second, to the extent that profit is weighted more heavily than consumers' surplus in political objective functions, the tendency for property rights to be too weak in equilibrium is offset. Finally, except in very special circumstances, equilibrium patent breadths are not identical in the two countries, and in all cases there is a continuum of optimal mixes of patent breadth, only one of which is symmetric. Further, if equilibrium breadths are sufficiently asymmetric, there is no symmetric patent policy that Pareto-dominates the original equilibrium. Like the main result of chapter 4, this last result directly applies to the current situation in Europe. Compared to chapter 4, however, the conclusion is completely different. Chapter 7 shows that the tendency in Europe towards unification is not necessarily (and only in exceptional cases) a Pareto improvement compared to the presence of different national patent systems.

Chapter 8 deals with the aim for unification at an even higher, global level. Through the past GATT Uruguay Round, northern countries have attempted to extend their strong standards of protection to southern, mainly underdeveloped, countries. Ever since the Paris Industrial Property Convention (1883), and more recently, the Washington Patent Cooperation Treaty (1970), there has been a movement toward uniform global protection. The GATT Round was used by northern countries to accelerate this process. Chapter 8 develops a simple locational model to study the question of optimal worldwide intellectual property protection. The world is represented as a line from north to south and each country is a point on the line. The image has been borrowed from the product differentiation literature discussed above. The north is defined as the group of countries that provide protection and the south as the group of countries that do not provide protection. Innovations only occur in the north, but they can be imitated (though imperfectly) in the south. The value of an innovation diminishes as the distance from the innovator becomes larger, either because international transport costs are incurred in exporting the product, or because the technology becomes less suitable for more distant countries. The analysis aims at determining the optimal point dividing north and south. Two aspects that are ignored in previous studies are central to this model: transport costs in international trade and imperfect imitation. It turns out that if the imitation capability in the south is poor, the globally optimal extent of intellectual property protection is worldwide protection. However, if southern imitations are better, it is optimal to have a southern region without effective intellectual property rights. The optimal world coverage also depends on the transport costs. If the transport costs to import a new product are high relative to the innovation cost to generate the new product, worldwide protection is, again, not optimal. These conclusions

are fairly robust and in general do not change for the various extensions of the basic model discussed.

Overall Conclusions

In addition to the conclusions emerging from the separate essays, some more general conclusions can be drawn from this book. The first deals with the research questions raised. The patent system has proven to be a powerful institution that has survived for centuries. This might be due to the simple idea behind it: in return for carrying out research and disclosing the fruits of research, provide temporary market power to the researcher. Although empirical evidence suggests that the patent system does not induce private firms to carry out R&D as strongly as often thought, it is a fact that a large number of firms do actually use the patent system. Yet, in spite of the considerable attention throughout the years, the economic theory has only recently recognized some obvious problems related to patents. The most outstanding example being the ever present practice of inventing around patents. Although this book contains some efforts in this direction, a lot of new questions remain to be raised and known questions need to be re-examined. Only after more research has been conducted do conclusions about the overall performance and desirability of the patent system seem valid.

The second general conclusion relates to the methodology used. The differentiation models applied in this book are not only good tools for the analysis of the economics of patent protection, but also for the economics of technical change in general. As Comanor (1967) already observed, much industrial R&D is spent in order to create differentiation. Through differentiation a firm can increase its market power and, consequently, its profit. Remarkably, the link between the economics of product differentiation and the economics of technical change is not better exploited. Certainly, there are fields that have recognized the relation. The new growth theory, for example, has implemented a micro basis which originates from (the non-address branch) of the product differentiation literature (see Grossman and Helpman 1991). Nevertheless, there remains much more to be exploited, particularly at the firm and industry level.

Future Research

On the basis of the studies performed in this book, several directions for future research can be given. I will only indicate three directions which, in my view, seem particularly promising and should be carried out first. First of all, the stage of R&D before patenting should be studied. Given the results in this book, patent races should not have a winner-takes-all but a losers-take-some structure. There are some examples of such models (for example, Mortensen 1982 and Stewart 1983), but more effort should be put in the development of models which incor-

porate a multiple-prize structure in the pre-patent stage on the one hand, and an imperfect protection structure which allows from inventing around in the post-patent stage on the other. The second direction concerns the dynamics of models of patent protection. Rather than the one basic innovation and one or two subsequent innovations studied here, future models should include more innovations in time. The dimension of patent height, for example, is richer than becomes clear in this book; the full effects of height are better visible if there is more than one patent in the market. The final direction is empirical research on the new patent dimensions. Chapter 6 has analysed one particular aspect, the proportion of patents that pass the patent office. A recent study by Lerner (1994) investigates the importance of the breadth of patent protection for the value of new firms in biotechnology. Given the availability of some highly detailed data-bases on patents, there are opportunities to examine other dimensions as well.

References

- Armington, P.S. 1969. "A Theory of Demand for Products Distinguished by Place of Production." *IMF Staff Papers* 16: 106-116.
- Anderson, S.P. and D.J. Neven. 1991. "Cournot Competition Yields Spatial Agglomeration." *International Economic Review* 32: 793-808.
- Anderson, S.P., A. de Palma, and J.-F. Thisse. 1992. *Discrete Choice Theory of Product Differentiation*. Cambridge MA: MIT Press.
- Archibald, G.C., B.C. Eaton and R.G. Lipsey. 1986. "Address Models of Value Theory." In J.E. Stiglitz and G.F. Mathewson (eds.), *New Developments in the Analysis of Market Structure*. London: MacMillan.
- Arrow, K.J. 1962. "Economic Welfare and the Allocation of Resources for Invention." In R.R. Nelson (ed.), *The Rate and Direction of Inventive Activity: Economic and Social Factors*. New York: Princeton University Press.
- Arrow, K.J. 1994. "The Production and Distribution of Knowledge." In G. Silverberg and L. Soete (eds.), *The Economics of Growth and Technical Change*. Aldershot: Edward Elgar Publishing.
- d'Aspremont, C., J. J. Gabszewicz and J.-F. Thisse. 1979. "On Hotelling's "Stability in Competition"." *Econometrica* 47: 1145-1151.
- Baker, R. 1976. *New and Improved (Inventors and Inventions that have Changed the Modern World)*. London: British Museum Publications Ltd.
- Barzel, Y. 1989. *Economic Analysis of Property Rights*. Cambridge: Cambridge University Press.
- Beath, J. and Y. Katsoulacos. 1991. *The Economic Theory of Product Differentiation*. Cambridge: Cambridge University Press.
- Beier, F.-K. and G. Schricker (eds). 1988. *GATT or WIPO? New Ways in the International Protection of Intellectual Property*. IIC Studies, Vol. 11.

- Berkowitz, M.K. and Y. Kotowitz. 1982. "Patent Policy in an Open Economy." *Canadian Journal of Economics* XV: 1-17.
- Besen, S.M. and L.J. Raskind. 1991. "An Introduction to the Law and Economics of Intellectual Property." *Journal of Economic Perspectives* 5: 3-27.
- de Bondt, R.R. 1977. "Innovative Activity and Barriers to Entry." *European Economic Review* 10: 95-109.
- Brinkhof, J.J. 1988. "Het Romeinse Recht en het Octrooirecht." *Bijblad Industriële Eigendom* 6: 122-124.
- de Brock, L.M. 1985. "Market Structure, Innovation and Optimal Patent Life." *Journal of Law and Economics* 4: 223-244.
- Buchanan, J.M. 1975. *The Limits of Liberty, between Anarchy and Leviathan*. Chicago: University of Chicago Press.
- van Cayseele, P. 1989. "On the Design of Patents for New Pharmaceuticals." *European Journal of Political Economy* 5: 49-61.
- Chin, J.C. and G.M. Grossman. 1990. "Intellectual Property Rights and North-South Trade." In R.W. Jones and A.O. Krueger (eds.), *The Political Economy of International Trade: Essays in Honor of Robert E. Baldwin*, Cambridge MA: Basil Blackwell.
- Chou, C.-F. and O. Shy. 1993. "The Crowding-Out Effects of Long Duration of Patents." *RAND Journal of Economics* 24 (2): 304-312.
- Clark, J., C. Freeman and L. Soete. 1983. "Long Waves, Inventions, and Innovations." In C. Freeman (ed.), *Long Waves in the World Economy*, London: Butterworth.
- Coase, R.H. 1960. "The Problem of Social Cost." *Journal of Law and Economics* 3: 1-44.
- Cohen, W.M. and D.A. Levinthal. 1989. "Innovation and Learning: The Two Faces of R&D." *Economic Journal* 99: 569-596.
- Comanor, W.S. 1967. "Market Structure, Product Differentiation, and Industrial Research." *Quarterly Journal of Economics* 81: 639-57.
- Cooper, C. (ed.) 1973. *Science, Technology and Development*. London: Frank Cass & Co.
- Cornes, R. and T. Sandler. 1986. *The Theory of Externalities, Public Goods, and Club Goods*. Cambridge: Cambridge University Press.
- Cornish, W.R. 1989. *Intellectual Property: Patents, Copyright, Trade Marks and Allied Rights*. London: Sweet & Maxwell.
- Crampes, C. and A. Hollander. 1991. "Duopoly and Quality Standards." Discussion Paper 91.21.236. GREMAQ, Université des Sciences Sociales. Toulouse.

Crampes, C. and M. Moreaux. 1993. "Les Caractéristiques des Brevets." Discussion Paper 93.09.296. GREMAQ, Université des Sciences Sociales. Toulouse.

Dasgupta, P. 1988. "Patents, Priority and Imitation or, The Economics of Races and Waiting Games." *Economic Journal* 98: 66-80.

Dasgupta, P. and E. Maskin. 1986. "The Existence of Equilibrium in Discontinuous Economic Games, I: Theory; and II: Applications." *Review of Economic Studies* LIII: 1-41.

David, P.A. 1991. "Technology, Resource Endowments, Property Rights and Trade: An Open Developing Country's Viewpoint." CEPR Discussion Paper No. 278, Stanford University.

David, P.A. 1993. "Intellectual Property Institutions and the Panda's Thumb: Patents, Copyrights, and Trade Secrets in Economic Theory and History." In M.B. Wallerstein, M.E. Mogee and R.A. Schoen (eds), *Global Dimensions of Intellectual Property Rights in Science and Technology*. Washington DC: National Academy Press.

David, P.A. and T.E. Olsen. 1992. "Technology Adoption, Learning Spillovers, and the Optimal Duration of Patent-Based Monopolies." *International Journal of Industrial Organization* 10: 517-543.

Deardorff, A.V. 1992. "Welfare Effects of Global Patent Protection." *Economica* 59: 35-51.

Debreu, G. 1959. *Theory of Value*. New Haven: Yale University Press.

van Dijk, T.W.P. 1992. "The Effects of Novelty Requirements on Subsequent Innovations." *Tinbergen Institute Research Bulletin* 4: 187-195.

van Dijk, T.W.P. 1993. "On the Exploitation of Patent Protection." Maastricht: MERIT Research Memorandum 93-010.

van Dijk, T.W.P. and P. van Cayseele. 1994. "Economic Implications of Converging Patent Breadth in Europe." In J. Hagedoorn (ed.), *Technical Change and the World Economy*. Cheltenham: Edward Elgar Publishing.

van Dijk, T.W.P. "Innovation Incentives through Third-Degree Price Discrimination in a Model of Patent Breadth." *Economics Letters*, forthcoming.

Diwan, I. and D. Rodrik. 1991. "Patents, Appropriate Technology, and North-South Trade." *Journal of International Economics* 30: 27-47.

Dixit, A.K. and J.E. Stiglitz. 1977. "Monopolistic Competition and Optimum Product Diversity." *American Economic Review* 67: 297-308.

Dosi, G. 1988. "Sources, Procedures and Microeconomic Effects of Innovation." *Journal of Economic Literature* 26: 1120-1171.

- Duysters, G. and J. Hagedoorn. 1994. "Convergence and Divergence in the International Information Technology." In J. Hagedoorn (ed.), *Technical Change and the World Economy*. Cheltenham: Edward Elgar Publishing.
- Eaton, B.C. and R.G. Lipsey. 1975. "The Principle of Minimum Differentiation Reconsidered: Some New Developments in the Theory of Spatial Competition." *Review of Economic Studies* 42: 27-49.
- Eaton, B.C. and R.G. Lipsey. 1978. "Freedom of Entry and the Existence of Pure Profit." *Economic Journal* 88: 455-469.
- Eaton, B.C. and R.G. Lipsey. 1989. "Product Differentiation." In R. Schmalensee and R.D. Willig (eds.), *Handbook of Industrial Organization*. Amsterdam: Elsevier.
- Economides, N. 1984. "The Principle of Minimum Differentiation Revisited." *European Economic Review* 24: 345-368.
- Economides, N. 1986. "Minimal and Maximal Product Differentiation in Hotelling's Duopoly." *Economics Letters* 21: 67-71.
- Economides, N. 1986. "Nash Equilibrium in Duopoly with Products defined by Two Characteristics." *RAND Journal of Economics* 17: 431-439.
- European Patent Convention. *Tractatenblad* 1975 (108). Den Haag: Staatsuitgeverij.
- Feinberg, R.M. and D.J. Rousslang. 1990. "The Economic Effects of Intellectual Property Right Infringements." *Journal of Business* 63: 79-90.
- Ferrantino, M.J. 1993. "The Effect of Intellectual Property Rights on International Trade and Development." *Weltwirtschaftliches Archiv* 129: 300-331.
- Freeman, C. 1982. *The Economics of Industrial Innovation*. London: Frances Pinter.
- Friden, G. 1989. "Recent Developments in EEC Intellectual Property Law: the Distinction between Existence and Exercise Revisited." *Common Market Law Review* 26: 193-217.
- Friedman, J.M. 1983. *Oligopoly Theory*. Cambridge: Cambridge University Press.
- Gallini, N.T. 1992. "Patent Policy and Costly Imitation." *RAND Journal of Economics* 23: 52-63.
- Gilbert, R. and C. Shapiro. 1990. "Optimal Patent Length and Breadth." *RAND Journal of Economics* 21: 106-112.
- Gilbert, R. and D.M.G. Newbery. 1982. "Preemptive Patenting and the Persistence of Monopoly." *American Economic Review* 72: 514-526.
- Goyal, S. and E.A.A. de Laat. 1992. "On the Optimality of Single vs. Sequential Patents." Mimeo. Erasmus University Rotterdam.

- Goyder, D.G. 1988. *EEC Competition Law*. Oxford: Clarendon Press.
- Griliches, Z. 1989. "Patents: Recent Trends and Puzzles." *Brooking Papers on Economic Activity* 5: 291-330.
- Griliches, Z. 1990. "Patent Statistics as Economic Indicators: A Survey." *Journal of Economic Literature* XXVIII: 1661-1707.
- Grossman, G.M. and E. Helpman. 1991. *Innovation and Growth in the Global Economy*. Cambridge MA: MIT Press.
- Grossman, G.M. and E. Helpman. 1994. "Endogenous Innovation in the Theory of Growth." *Journal of Economic Perspectives* 8: 23-44.
- Grossman, G.M. and C. Shapiro. 1987. "Dynamic R&D Competition." *Economic Journal* 97: 372-387.
- Grossman, G.M. and C. Shapiro. 1988. "Counterfeit-Product Trade." *American Economic Review* 78: 59-75.
- Hall, B.H., Griliches, Z. and J.A. Hausman. 1986. "Patents and R&D: Is There a Lag?" *International Economic Review* 27: 265-83.
- Harris, C. and J. Vickers. 1987. "Racing with Uncertainty." *Review of Economic Studies* LIV: 1-21.
- Hausman, J.A. and J.K. MacKie-Mason. 1988. "Price Discrimination and Patent Policy." *RAND Journal of Economics* 19: 253-265.
- Helpman, E. 1993. "Innovation, Imitation, and Intellectual Property Rights." *Econometrica* 61(6): 1247-1280.
- Hirshleifer, J. 1971. "The Private and Social Value of Information and the Reward to Inventive Activity." *American Economic Review* 61: 561-574.
- Hoffmann, W.D. 1976. "Market Structure and Strategies of R&D Behaviour in the Data Processing Market - Theoretical Thoughts and Empirical Findings." *Research Policy* 5: 334-353.
- Horstmann, I., G.M. MacDonald and A. Slivinski. 1985. "Patents as Information Transfer Mechanisms: To Patent or (Maybe) Not to Patent." *Journal of Political Economy* 93: 837-858.
- Hotelling, H. 1929. "Stability in Competition." *Economic Journal* 39: 41-57.
- Hoyng, W.A. 1990. *Beschermingsomvang Anno 1989*. Deventer: Kluwer.
- Ireland, N.J. 1987. *Product Differentiation and Non-Price Competition*. Oxford: Basil Blackwell Ltd.

- Judd, K.L. 1985. "On the Performance of Patents." *Econometrica* 53: 567-586.
- Kaldor, N. 1935. "Market Imperfection and Excess Capacity." *Economica* 2: 33-50. Reprinted in: Kaldor, N. 1980. *Essays on Value and Distribution*. London: Duckworth.
- Kaldor, N. 1957. "A Model of Economic Growth." *Economic Journal*. Reprinted in: Kaldor, N. 1980. *Essays on Economic Stability and Growth*. London: Duckworth.
- Kamien, M.I. 1992. "Patent Licensing." In R.J. Aumann and S. Hart (eds.), *Handbook of Game Theory with Economic Applications*. Amsterdam: Elsevier.
- Kamien, M.I. and N.L. Schwartz. 1974. "Patent Life and R&D Rivalry." *American Economic Review* 64: 183-187.
- Kamien, M.I. and N.L. Schwartz. 1982. *Market Structure and Innovation*. Cambridge: Cambridge University Press.
- Kaplow, L. 1984. "The Patent-Antitrust Intersection: A Reappraisal." *Harvard Law Review* 97: 1813-1892.
- Kaufer, E. 1989. *The Economics of the Patent System*. London: Harwood Academic Publishers.
- Kitch, E.W. 1977. "The Nature and Function of the Patent System." *Journal of Law and Economics* 20: 265-290.
- Klemperer, P. 1990. "How Broad should the Scope of Patent Protection Be?" *RAND Journal of Economics* 21: 113-130.
- Kunz-Hallstein, H.P. 1989. "The U.S. Proposal for a GATT-Agreement on Intellectual Property and the Paris Convention for the Protection of Industrial Property." In F.-K. Beier and G. Schricker (eds). *GATT or WIPO? New Ways in the International Protection of Intellectual Property*. IIC Studies, Vol. 11.
- Krugman, P. 1979a. "A Model of Innovation, Technology Transfer, and the World Distribution of Income." *Journal of Political Economy* 87: 253-266.
- Krugman, P. 1979b. "Increasing Returns, Monopolistic Competition and International Trade." *Journal of International Economics* 9: 469-479.
- Lancaster, K.J. 1980. "Intra-Industry Trade under Perfect Monopolistic Competition." *Journal of International Economics* 10: 151-175.
- Leamer, E. and R. Stern. 1970. *Quantitative International Economics*. Boston.
- Levin, R.C., A.K. Klevorick, R.R. Nelson, and S.G. Winter. 1987. "Appropriating the Returns from Industrial Research and Development." *Brookings Papers on Economic Activity* 3: 783-820.

- Lerner, J. 1994. "The Importance of Patent Scope: An Empirical Analysis." *RAND Journal of Economics* 25: 319-333.
- Linneman, H. 1966. *An Econometric Study of International Trade Flows*. Amsterdam.
- Lösch, A. 1940. *The Economics of Location*. Jena: Fischer. English translation. New Haven: Yale University Press, 1954.
- Loury, G.L. 1979. "Market Structure and Innovation." *Quarterly Journal of Economics* XCIII: 395-410.
- Machlup, F. and E. Penrose. 1950. "The Patent Controversy in the Nineteenth Century." *Journal of Economic History* 10: 1-29.
- Machlup, F. 1958. *An Economic Review of the Patent System*. Study no. 15 of the Subcommittee on Patents, Trademarks and Copyrights of the Committee on the Judiciary, U.S. Senate. Washington DC: Government Printing Office.
- Maddala, G.S. 1983. *Limited-dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.
- la Manna, M. 1992. "Optimal Patent Life vs. Optimal Patentability Standards." *International Journal of Industrial Organization* 10: 81-90.
- la Manna, M. 1992. "New Dimensions of the Patent System." In G. Norman and M. la Manna (eds.), *The New Industrial Economics, Recent Developments in Industrial Organization, Oligopoly and Game Theory*. Aldershot: Edward Elgar.
- la Manna, M., R. MacLeod, and D. de Meza. 1989. "The Case for Permissive Patents." *European Economic Review* 33: 1427-1443.
- Mansfield, E., M. Schwartz, and S. Wagner. 1981. "Imitation Costs and Patents: An Empirical Study." *Economic Journal* 91: 907-918.
- Mansfield, E. 1986. "Patents and Innovation." *Management Science* 32: 173-181.
- Margerison. 1978. "Computers." In T.I. Williams (ed.), *A History of Technology* (vol. VII part II). Oxford: Clarendon Press.
- Matutes, C., P. Regibeau, and K. Rockett. 1993. "Optimal Patent Design and Diffusion of Innovations." Mimeo. Northwestern University. Revised version.
- Merges, R.P. and R.R. Nelson. 1992. "Market Structure and Technical Advance: The Role of Patent Scope Decisions." In T.M. Jorde and D.J. Teece (eds.), *Antitrust, Innovation, and Competitiveness*. Oxford: Oxford University Press.
- Meurer, M.J. 1989. "The Settlement of Patent Litigation." *RAND Journal of Economics* 20: 77-91.

- de Meza, D. and J.R. Gould. 1992. "The Social Efficiency of Private Decisions to Enforce Property Rights." *Journal of Political Economy* 100: 561-580.
- Mortensen, D.T. 1982. "Property Rights and Efficiency in Mating, Racing and Related Games." *American Economic Review* 72: 968-979.
- Nelson, R.R. 1959. "The Simple Economics of Basic Scientific Research." *Journal of Political Economy* 67: 297-306.
- Neven, D. and J.F. Thisse. 1990. "On Quality and Variety Competition." In J.J. Gabszewicz, J.-F. Richard and L.A. Wolsey (eds.), *Economic Decision-Making: Games, Econometrics and Optimisation*. Amsterdam: North-Holland.
- Nordhaus, W.D. 1969. *Invention, Growth, and Welfare, A Theoretical Treatment of Technological Change*. Cambridge MA: MIT Press.
- Ordover, J.A. 1991. "A Patent System for Both Diffusion and Exclusion." *Journal of Economic Perspectives* 5:43-60.
- Pakes, A. 1986. "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks." *Econometrica* 54: 755-784.
- de Palma, A., V. Ginsburgh, Y.Y. Papageorgiou and J.-F. Thisse. 1985. "The Principle of Minimum Differentiation holds under Sufficient Heterogeneity." *Econometrica* 53: 767-781.
- Papoulis, A. 1984. *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill.
- Penrose, E.T. 1951. *The Economics of the International Patent System*. Baltimore: John Hopkins Press.
- Philips, L. 1989. *The Economics of Price Discrimination*. Cambridge: Cambridge University Press.
- Pieroen, A.P. 1988. *Beschermingsomvang van Octrooien in Nederland, Duitsland en Engeland (Scope of Protection of Patents in the Netherlands, the Federal Republic of Germany and the United Kingdom)*. Deventer: Kluwer.
- Reinganum, J.F. 1989. "The Timing of Innovation: Research, Development, and Diffusion." In R. Schmalensee and R.D. Willig (eds.), *Handbook of Industrial Organization*. Amsterdam: Elsevier.
- Romer, P. 1993. "Implementing a National Technology Strategy with Self-Organizing Industry Investment Boards." *Brookings Papers on Economic Activity* 2: 345-399.
- Romer, P. 1994. "The Origin of Endogenous Growth." *Journal of Economic Perspectives* 8: 3-22.

- Ronnen, U. 1991. "Minimum Quality Standards, Fixed Costs, and Competition." *RAND Journal of Economics* 22: 490-504.
- Salop, S. 1979. "Monopolistic Competition with Outside Goods." *RAND Journal of Economics* 10: 141-156.
- Schankerman, M. and A. Pakes. 1986. "Estimates of the Value of Patent Rights in European Countries during the Post-1950 Period." *Economic Journal* 96: 1052-1076.
- Scherer, F.M. 1983. "The Propensity to Patent." *International Journal of Industrial Organisation* 1: 107-128.
- Scherer, F.M. and D. Ross. 1990. *Industrial Market Structure and Economic Performance*. Boston: Houghton Mifflin Company.
- Schiff, E. 1971. *Industrialization Without National Patents: The Netherlands 1869-1912, Switzerland 1850-1907*. Princeton: Princeton University Press.
- Schmalensee, R. 1981. "Output and Welfare Implications of Monopolistic Third-Degree Price Discrimination." *American Economic Review* 71: 242-247.
- Schumpeter, J. 1975. *Capitalism, Socialism and Democracy*. New York: Harper & Row.
- Scotchmer, S. 1991. "Standing on the Shoulder of Giants: Cumulative Research and the Patent Law." *Journal of Economic Perspectives* 5: 29-41.
- Scotchmer, S. and J. Green. 1990. "Novelty and Disclosure in Patent Law", *RAND Journal of Economics* 21: 131-146.
- Shaked, A. and J. Sutton. 1982. "Relaxing Price Competition through Product Differentiation." *Review of Economic Studies* 49: 3-13.
- Siegel, S. and N.J. Castellan, Jr. 1988. *Nonparametric Statistics*. New York: McGraw-Hill.
- Soete, L. 1981. "A General Test of Technology Gap Trade Theory." *Weltwirtschaftliches Archiv* 117: 638-666.
- Soete, L. and M. Cimoli. 1992. "A Generalized Technology Gap Trade Model." *Economie Appliquée* 3: 33-54.
- Solow, R.M. 1957. "Technical Progress and the Aggregate Production Function." *Review of Economics and Statistics* 39: 312-320.
- Spence, A.M. 1976. "Product Selection, Fixed Costs, and Monopolistic Competition." *Review of Economic Studies* 43: 217-235.
- Steurs, G. 1993. "Inter-Industry R&D Spill-Over: What Difference Do They Make?" Mimeo. Department of Applied Economics, Catholic University Leuven.

- Stewart, M.B. 1983. "Noncooperative Oligopoly and Preemptive Innovation Without Winner-Take-All." *Quarterly Journal of Economics* 98: 681-694.
- Tandon, P. 1982. "Optimal Patents with Compulsory Licensing." *Journal of Political Economy* 90: 470-486.
- Taylor, M.S. 1993. "TRIPs, Trade, and Technology Transfer." *Canadian Journal of Economics* XXVI: 625-637.
- Taylor, C.T. and Z.A. Silberston, 1973. *The Economic Impact of the Patent System: A Study of the British Experience*. Cambridge: Cambridge University Press.
- Tinbergen, J. 1942. "Zur Theorie der Langfristigen Wirtschaftsentwicklung." *Weltwirtschaftliches Archiv*.
- Tirole, J. 1989. *The Theory of Industrial Organization*. Cambridge MA: MIT Press.
- Trajtenberg, M. 1990. "A Penny for Your Quotes: Patent Citations and the Value of Innovations." *RAND Journal of Economics* 21: 172-187.
- Ullrich, H. 1989. "GATT: Industrial Property Protection, Fair Trade and Development." In F.-K. Beier and G. Schricker (eds.), *GATT or WIPO? New Ways in the International Protection of Intellectual Property*. IIC Studies, Vol. 11.
- UNCTAD. 1989. *Uruguay Round: Papers on Selected Issues*. New York: United Nations.
- Vanhaverbeke, W. and P. van Cayseele. 1992. "On the Differential Effects of European Patents." *Economics of Innovation and New Technologies* 2: 185-201.
- Varian, H.L. 1985. "Price Discrimination and Social Welfare." *American Economic Review* 75: 870-875.
- Varian, H.L. 1989. "Price Discrimination." In R. Schmalensee and R.D. Willig (eds.), *Handbook of Industrial Organization*. Amsterdam: Elsevier.
- Vernon, R. 1957. *The International Patent System and Foreign Policy*. Study no. 5 of the Subcommittee on Patents, Trademarks and Copyrights of the Committee on the Judiciary, U.S. Senate. Washington DC: Government Printing Office.
- Verspagen, B. 1991. "A New Empirical Approach to Catching Up or Falling Behind." *Structural Change and Economic Dynamics* 2: 359-380.
- Waterson, M. 1990. "The Economics of Product Patents." *American Economic Review* 80: 860-869.
- White, T.A.B., R. Jacob, and J.D. Davies. 1978. *Patents, Trade Marks, Copyright and Industrial Designs*. London: Sweet & Maxwell.

van Wijk, J., J.I. Cohen and J. Komen. 1993. "Intellectual Property Rights for Agricultural Biotechnology: Options and Implications for Developing Countries." ISNAR Research Report 3, October.

(Summary in Dutch)

Williams, M.R. 1985. *A History of Computing Technology*. Englewood Cliffs: Prentice-Hall

Wright, B.D. 1983. "The Economics of Invention Incentives: Patents, Prizes, and Research Contracts." *American Economic Review* 73: 691-709.

Dit proefschrift gaat over de beperking van octrooibescherming. Een octrooi¹ is een tijdelijk en sterk exclusief eigendomsrecht op een uitvinding. Niet alle uitvindingen kunnen geoctrooid worden; er moet voldaan zijn aan bepaalde voorwaarden om voor octrooibescherming in aanmerking te komen. Zo moet een uitvinding ten eerste voldoende nieuw zijn ten opzichte van bestaande producten of processen, ten tweede niet voor de hand liggen voor experts in een bepaald technisch gebied, en tenslotte industrieel toepasbaar zijn. Voordat een uitvinding aan deze ten nog een aantal andere) eisen en krijgt daarmee de octrooiaanvrage door de beoordelingsprocedure van het octrooibureau bevestigd, dan wordt een octrooi verleend en is de uitvinding beschermd.

Wat houdt deze octrooibescherming precies in? Het thema van dit boek is de beperking van octrooibescherming. De bescherming die een octrooi biedt is niet, zoals vaak werd gedacht in de economische wetenschap, volledig maar er zijn diverse grenzen. In dit boek worden verschillende grenzen van octrooibescherming vastgesteld en bestudeerd. Het hoofdstuk gaat in op een bepaalde grens. Alle analyses beperken zich tot nieuwe of verbeterde producten; proces innovaties worden buiten beschouwing gelaten.

Stel dat een bedrijf een nieuw produkt heeft voortgebracht met behulp van omvangrijke investeringen in Onderzoek en Ontwikkeling (O&O). Stel verder dat dit bedrijf een octrooi krijgt op deze produkt innovatie. De meeste octrooien zijn tegenwoordig trouwens in handen van bedrijven, en niet van individuele uitvinders. Moderne ondernemingen hebben namelijk vaak O&O afdelingen waar voortdurend naar nieuwe, of verbeteringen van bestaande, producten wordt gezocht. Dit intentionele zoeken naar innovaties waarmee kosten gepaard gaan, vormt het belangrijkste economische argument voor het instellen van octrooien. Octrooien stimuleren investeringen in O&O doordat ze bedrijven beter in staat

1. Dierlijk of Nederlands voor het Engelse 'patent'.

Samenvatting (Summary in Dutch)

Dit proefschrift gaat over de beperking van octrooibescherming. Een octrooi¹ is een tijdelijk en sterk exclusief eigendomsrecht op een uitvinding. Niet alle uitvindingen kunnen geoctrooieerd worden; er moet voldaan zijn aan bepaalde voorwaarden om voor octrooibescherming in aanmerking te komen. Zo moet een uitvinding ten eerste voldoende nieuw zijn ten opzichte van bestaande produkten of processen, ten tweede niet voor de hand liggen voor experts in een bepaald technisch gebied, en tenslotte industrieel toepasbaar zijn. Voldoet een uitvinding aan deze (en nog een aantal andere) eisen en komt daarmee de octrooiaanvraag door de beoordelingsprocedure van het octrooibureau heen, dan wordt een octrooi verleend en is de uitvinding beschermd.

Wat houdt deze octrooibescherming precies in? Het thema van dit boek is de beperking van octrooibescherming. De bescherming die een octrooi biedt is niet, zoals vaak werd gedacht in de economische wetenschap, volledig maar er zijn diverse grenzen. In dit boek worden verschillende grenzen van octrooibescherming vastgesteld en bestudeerd. Ieder hoofdstuk gaat in op een bepaalde grens. Alle analyses beperken zich tot nieuwe of verbeterde produkten; proces innovaties worden buiten beschouwing gelaten.

Stel dat een bedrijf een nieuw produkt heeft voortgebracht met behulp van omvangrijke investeringen in Onderzoek en Ontwikkeling (O&O). Stel verder dat dit bedrijf een octrooi krijgt op deze produkt innovatie. De meeste octrooien zijn tegenwoordig trouwens in handen van bedrijven, en niet van individuele uitvinders. Moderne ondernemingen hebben namelijk vaak O&O afdelingen waar voortdurend naar nieuwe, of verbeteringen van bestaande, produkten wordt gezocht. Dit intentionele zoeken naar innovaties waarmee kosten gepaard gaan, vormt het belangrijkste economische argument voor het instellen van octrooien. Octrooien stimuleren investeringen in O&O doordat ze bedrijven beter in staat

1. Octrooi is Nederlands voor het Engelse 'patent'.

stellen om deze investeringen terug te verdienen. Als nieuwe produkten ook gegenereerd zouden worden zonder investeringen in O&O, dan zou er geen economisch motief zijn voor het instellen van octrooien. Bedrijven dienen gestimuleerd te worden in onderzoek omdat de sociale opbrengsten (dat is, voor de gemeenschap als geheel) van de innovaties die eruit voortkomen vaak veel groter zijn dan de private opbrengsten voor een bedrijf alleen. Zonder octrooien zouden bedrijven dus te weinig in innovaties investeren.

Wat betekent octrooibescherming voor een bedrijf? Hoofdstuk 3 bekijkt de dreiging die uitgaat van concurrenten die met verbeteringen van een geoctrooieerd produkt komen. Stel dat de octrooihouder één concurrent heeft. Bij introductie van het nieuwe produkt op de markt leert deze concurrent het nieuwe produkt kennen. Echter, vanwege het octrooi is het deze concurrent niet toegestaan om exact hetzelfde produkt ook op de markt te brengen. Wat de concurrent wel kan doen is proberen 'om het octrooi heen uit te vinden'. Hij kan bijvoorbeeld zijn O&O afdeling opdracht geven om naar verbeteringen van het produkt te zoeken. In geval van een nieuw geneesmiddel bijvoorbeeld kan worden gezocht naar nieuwe samenstellingen die minder negatieve bijwerkingen veroorzaken. Als de verbetering voldoende groot is ten opzichte van het bestaande octrooi, wordt aan de nieuwigheidseisen van het octrooibureau voldaan. Het is de concurrent dan toegestaan om met zijn verbetering de markt te betreden. Het hangt dus van de strengheid van beoordelen van het octrooibureau af hoe groot de dreiging van verbeteringen voor de octrooihouder is. Hoofdstuk 3 introduceert hiervoor de 'hoogte' van octrooibescherming, gegeven door de strengheid van beoordelen. Hoe hoger de bescherming voor de octrooihouder - dus hoe strenger de beoordeling in het octrooibureau -, hoe moeilijker het is voor een concurrent om een voldoende verbetering te genereren en tot de markt toe te treden. De effecten van octrooihoogte op de concurrentie in een markt kunnen in drie categorieën ingedeeld worden. In de eerste categorie van lage octrooibescherming hebben octrooien geen invloed op de natuurlijke marktuitsluiting (zonder octrooien). De reden hiervoor is dat een concurrent met zijn verbetering een natuurlijke afstand houdt van het geoctrooieerde produkt met als doel de prijsconcurrentie minder sterk te maken. In de middencategorie kan een octrooi meer winst opleveren voor de octrooihouder en minder voor de concurrent, omdat deze een grotere verbetering moet genereren dan optimaal voor hem is. In deze categorie kan het hebben van een octrooi echter ook een handicap zijn, omdat het verplicht een minder winstgevende positie te kiezen. De laatste categorie, tenslotte, van hoge octrooibescherming, is altijd voordelig voor de octrooihouder omdat het octrooi hem dan een absolute monopoliepositie verschaft.

In hoofdstuk 4 wordt bekeken welke restricties een octrooihouder kan onder-

vinden door regels van mededingingsbeleid. Er bestaat een bijna natuurlijk conflict tussen octrooi- en mededingingsbeleid; octrooi-beleid verschaft monopolieposities (om innovaties te stimuleren), terwijl mededingingsbeleid vaak monopolieposities bestrijdt (om de sociale kosten die worden veroorzaakt door monopolies te beperken). Het hoofdstuk spitst zich toe op de 'breedte' van octrooi-bescherming en de praktijk van prijsdiscriminatie.

Ik zal eerst op de octrooi-breedte ingaan. Zoals eerder gezegd is octrooi-hoogte een indicatie van de bescherming tegen *verbeteringen* van een geoctrooieerd produkt. Octrooi-breedte daarentegen geeft de mate van bescherming aan tegen *imitaties* van het nieuwe produkt. Wat is het verschil tussen hoogte en breedte, en daarmee samenhangend, tussen verbeteringen en imitaties? Ten eerste is er een technisch verschil. Hoogte geeft het minimum aantal nieuwe elementen aan dat een verbetering moet bevatten om door het octrooibureau heen te geraken. Breedte geeft het aantal elementen van het geoctrooieerde produkt aan dat een gelijkaardig produkt (een imitatie) maximaal mag bevatten om geen inbreuk te plegen op het octrooi. Dus hoogte is een indicatie van de vereiste nieuwheid en breedte van de toegestane gelijkheid. Een tweede verschil betreft de kosten van verbeteringen en imitaties. Verbeteringen vereisen over het algemeen meer investeringen in O&O, omdat nieuwe elementen moeten worden voortgebracht, dan imitaties, waarin bestaande elementen worden nagemaakt. Een derde verschil tenslotte heeft te maken met de perceptie van consumenten. Tegen gelijke prijzen zal een consument bijvoorbeeld een verbeterd geneesmiddel met minder bijwerkingen prefereren boven het oudere geneesmiddel. In het geval van imitaties zullen sommige consumenten het originele middel prefereren terwijl andere consumenten geïmiteerde, gelijkaardige varianten prefereren.

Omdat octrooi-breedte de bescherming tegen gelijkaardige substituten aan geeft, is het in essentie een maat van monopolie-macht van de octrooihouder. Als nu de breedte die een octrooihouder wordt verschaft verschilt tussen twee landen dan zou, *ceteris paribus*, de monopolie-macht en dus de optimale prijs per land verschillen. Echter, vaak wordt het vanuit mededingingsbeleid verboden om verschillende prijzen voor hetzelfde produkt te vragen. In Europa bijvoorbeeld is prijsdiscriminatie niet toegestaan. De reden hiervoor is dat, onder bepaalde voorwaarden, de welvaart bij prijsdiscriminatie lager is dan bij uniforme prijzen. De winst echter van een bedrijf dat prijsdiscriminatie toepast, is over het algemeen hoger dan van een bedrijf dat uniforme prijzen hanteert. Hoofdstuk 4 gaat in op de vraag hoe een potentiële innovator het best, vanuit sociaal oogpunt, een bepaalde minimum winst, nodig om hem tot O&O aan te zetten, kan worden voorzien. Zoals gezegd levert prijsdiscriminatie meer winst op maar minder welvaart, terwijl uniforme prijzen minder winst maar meer welvaart opleveren. De vraag is dus wat de meeste welvaart *per eenheid winst* oplevert. In een model

Curriculum Vitae

Theon van Dijk werd geboren op 6 december 1966 te Oss. In 1985 sloot hij met goed gevolg het Gymnasium- α af aan het Titus Brandsma Lyceum te Oss. Van 1985 to 1989 studeerde hij algemene economie aan de Rijksuniversiteit Limburg te Maastricht. Het doctoraal diploma werd cum laude behaald. Van 1989 tot 1994 was hij, aanvankelijk als Assistent-In-Opleiding (AIO) en later als Onderzoeker-In-Opleiding (OIO), verbonden aan het Maastricht Economic Research Institute on Innovation and Technology (MERIT), onderdeel van de Faculteit der Economische Wetenschappen, Rijksuniversiteit Limburg. Momenteel is hij als Postdoctoral Research Fellow verbonden aan ditzelfde onderzoeksinstituut. Vanaf 1 oktober 1994 zal hij als Postdoctoral Research Fellow werkzaam zijn aan het Institut d'Économie Industrielle (IDEI), Université des Sciences Sociales, Toulouse, Frankrijk.

THE LIMITS OF PATENT PROTECTION

Essays on the Economics of Intellectual Property Rights

Theon van Dijk

This book investigates aspects of patent protection which have been neglected in economic science so far. When observing patent laws, procedures, court cases and business practice, various limits of patent protection become apparent. These limits form the common theme of this book. Chapter 3 studies the threat for a patentholder stemming from competitors who enter the market with improvements of the patented innovation. Chapter 4 examines the restrictions imposed upon a patentholder by rules of competition policy. Next, chapter 5 investigates strategies pursued by firms that invent around a patent in a legal way. The empirical study of chapter 6 reveals how firms in the data-processing industry have coped with examiners in the European Patent Office. The international dimension enters the analysis in the last part of the book. Chapter 7 explains why patent protection in one country can differ from that offered in another country. Chapter 8 examines the question of how far patent protection should be extended worldwide. This issue has been high on the agenda for the negotiations between the Northern and the Southern countries in the Gatt Uruguay Round.

Datawyse | Universitaire Pers Maastricht
ISBN 90 5278 153 2



Faculty of Economics and Business Administration
University of Limburg
Maastricht, the Netherlands
Dissertation no. 64-21